Some Essays on models in the Bond and Energy Markets

Oleksandr Castello

Department of Economics
University of Genoa

This dissertation is submitted in conformity with the requirements for the degree of *Doctor of Philosophy* (XXXV Cycle)

June 2023

Acknowledgements

Firstly, I would like to express my special thanks to my supervisor, professor Marina Resta, who inspired me to undertake this academic path. Especially i am grateful for the help, continuous support and for guiding me through the steps of conducting my Ph.D research and in writing this dissertation. I could not have imagined having better academic mentor.

My sincere thanks also goes to professor Anna Bottasso for the support in the nascent stages of my doctoral studies.

Finally, I want to wholeheartedly thank my parents for their love, patience, constant encouragement and endless support that has enabled me to pursue this path. They are the pillars of my life, my safe harbour, and i will always be grateful to them for everything.

Abstract

The term structure of interest rates plays a fundamental role as an indicator of economy and market trends, as well as a supporting tool for macroeconomic strategies, investment choices or hedging practices. Therefore, the availability of proper techniques to model and predict its dynamics is of crucial importance for players in the financial markets.

Along this path, the dissertation initially examined the reliability of parametric and neural network models to fit and predict the term structure of interest rates in emerging markets, focusing on the Brazilian, Russian, Indian, Chines and South African (BRICS) bond markets. The focus on the BRICS is straightforward: the dynamics of their term structures make tricky the application of consolidated yield curve models. In this respect, BRICS yield curve act as stress testers.

The study then examined how to apply the above cited models to energy derivatives, focusing the attention on the Natural Gas and Electricity futures, motivated by the existence of similarity. The research was carried out using ad hoc routines, such as the R package "DeRezende.Ferreira", developed by the candidate and now freely downloadable at the Comprehensive R Archive Network (CRAN) repository[1], as well as by means of code written in MatLab 2021a - 2022a and Python (3.10.10) using the open-source Keras (2.4.3) library with TensorFlow (2.4.0) as backend.

The dissertation consists of four chapters based on published and/or under submission materials. Chapter 1 is an excerpt of the paper

- Castello, O.; Resta, M. Modeling the Yield Curve of BRICS Countries: Parametric vs. Machine Learning Techniques. Risks 2022

The work firstly offers a comprehensive analysis of the BRICS bond market and then investigates and compares the abilities of the parametric Five–Factor De Rezende–Ferreira model and Feed–Forward Neural Networks to fit the yield curves. Chapter 2 is again focused on the BRICS market but investigates a methodology to identify optimal

[1] https://cran.r-project.org/web/packages/DeRezende.Ferreira/index.html

time–varying parameters for parametric yield curve models. The work then investigates the ability of this method both for in–sample fitting and out–of–sample prediction. Various forecasting methods are examined: the Univariate Autoregressive process AR(1), the TBATS and the Autoregressive Integrated Moving Average (ARIMA) combined to Nonlinear Autoregressive Neural Networks (NAR–NN). Chapter 3 studies the term structure dynamics in the Natural Gas futures market. This chapter represents an extension of the paper

- Castello, O., Resta, M. (2022). Modeling and Forecasting Natural Gas Futures Prices Dynamics: An Integrated Approach. In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds) Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2022.

After showing that the natural gas and bond markets share similar stylized facts, we exploit these findings to examine whether techniques conventionally employed on the bonds market can be effectively used also for accurate in–sample fitting and out–of–sample forecast. We worked at first in–sample and we compared the performance of three models: the Four–Factor Dynamic Nelson–Siegel–Svensson (4F-DNSS), the Five–Factor Dynamic De Rezende–Ferreira (5F–DRF) and the B–Spline. Then, we turned the attention on forecasting, and explored the effectiveness of a hybrid methodology relying on the joint use of 4F–DNSS, 5F–DRF and B–Splines with Nonlinear Autoregressive Neural Networks (NAR–NNs). Empirical study was carried on using the Dutch Title Transfer Facility (TTF) daily futures prices in the period from January 2011 to June 2022 which included also recent market turmoil to validate the overall effectiveness of the framework.

Chapter 4 analyzes the predictability of the electricity futures prices term structure with Artificial Neural Networks. Prices time series and futures curves are characterized by high volatility which is a direct consequence of an inelastic demand and of the non–storable nature of the underlying commodity. We analyzed the forecasting power of several neural network models, including Nonlinear Autoregressive (NAR–NNs), NAR with Exogenous Inputs (NARX–NNs), Long Short–Term Memory (LSTM–NNs) and Encoder–Decoder Long Short–Term Memory Neural Networks (ED–LSTM–NNs). We carried out an extensive study of the models predictive capabilities using both the univariate and multivariate setting. Additionally, we explored whether incorporating various exogenous components such as Carbon Emission Certificates (CO_2) spot prices, as well as Natural Gas and Coal futures prices can lead to improvements of the models performances. The data of the European Energy Exchange (EEX) power market were adopted to test the models. Chapter 4 concludes.

Contents

List of Figures

List of Tables

Chapter 1

Modeling the Yield Curve of BRICS Countries: Parametric vs. Machine Learning Techniques

1.1 Introduction

The term structure of interest rates, whose graphical representation is given by the yield curve, describes the relationship between market interest rates and different times to maturity, and provides an ex–ante measure of the investor's return in a fixed income market (Saunders and Cornett, 2014). Besides, the yield curve contains fundamental information to analyze the economic and financial situation of a country, which can be interpreted in terms of market expectations of monetary policy, economic activity and inflation over short, medium and long–term horizons; for this reason it is often employed to support macroeconomic strategies. Modeling it is therefore fundamental for financial economists and risk managers to define hedging and pricing strategies, as well as to get an effective assessment of portfolio risk (Pereda, 2009). Furthermore, yield curves can also provide valuable information as input for financial stability, and banking supervision. Besides, once a nominal yield curve is computed, a term structure of real interest rates and break-even inflation rates can be derived.

When estimating yield curves, an important challenge is that they should reflect as many as possible relevant movements in the underlying term structure of interest rates. In the past decades an extensive literature has been developed accordingly: models based on stochastic processes (Hess, 2020), methods relying on splines (Filipović, 2009), factor models (Diebold and Rudenbusch, 2017) and techniques based on machine

learning algorithms (Lopez De Prado, 2018), to cite more relevant research strands. Shifting our attention to practical applications, Chakroun and Abid (2014), Ullah and Bari (2018) pinpointed that the great majority of the works focuses on developed countries, mostly the United States and the Eurozone and, on the contrary, relatively lower attention has been paid to emergent markets, despite their growing economic and political weight.

This (relative) lack of contributions on emerging economies is potentially troublesome, as the yield curve of those countries usually exhibits a very volatile behaviour with frequent and marked humps, in contrast to more developed countries whose yield curves are less sensitive. This rationale inspired our study, focused on the group of countries referred by the acronym BRICS (Brazil, Russia, India, China and, since 2010, South Africa) which are actually under the magnifying lens of financial investors as drivers of the globalization process of financial markets (Stuart, 2020). The BRICS countries, in fact, have experienced several years of rapid expansion in trade and economic growth: they currently account for nearly a quarter of the world economy. Furthermore, BRICS have set up the New Development Bank (NDB) where those countries address the group's economic challenges with combined resources. All these elements make the BRICS countries important players in the current pattern of global investment, because they are both the major recipients of foreign direct investments and increasingly important for outward investors. This, in turn, can have deep impact on the exposure to country risk: the economic, political and social contexts can eventually cause losses to foreign investors. After all, BRICS countries inspired many research strands Bekiros and Avdoulas (2020), de Boyrie and Pavlova (2016), Zeb and Rashid (2019), Salisu et al. (2021); however, with the exception of El-Shagi and Jiang (2019) and Caldeira et al. (2020), to the best of our knowledge, there aren't other works analyzing altogether BRICS yield curves and at the same time investigating on them the interpolation capabilities of various techniques. Such an investigation can be useful under various viewpoints: many central banks use to interpolate yield curves to assess monetary policy measures; in addition such models have an economic interpretation and they can be useful for measuring risk in fixed income portfolios. With this motivation, our paper tried to test whether modeling based on either the parametric or machine learning approach can ensure a streamlined in–sample fitting also for these new world players. To such aim, we compared two alternative methods, the Dynamic Five Factor parametric model (5F–DRF) (De Rezende and Ferreira, 2013) and the multilayer Feed–Forward Neural Network (FFNNs). The models were chosen among the most performing ones (both parametric and non–parametric). Focusing on the in–sample

rather than on the out–of–sample fit was supported by the existing literature, see for instance (R.R. Wahlstrom et al., 2021), and (Prasanna and Sowmya, 2017) that specifically refer to in–sample fitting of the yield curve. Moreover, as our work examines for the first time the yield curve of all the BRICS at once, it cannot be granted that methods of consolidated use for other countries can adapt so well also to our data, especially in presence of sudden spikes. Working in–sample is therefore nothing but the first step to identify whether those models work well also on emerging countries and which of them is the best model in this task. Answering to those questions, in turn, will allow to better address the forecasting issue for future research.

The 5F–DRF belongs to the factor models family (Diebold and Rudenbusch, 2017) which through the past decade have gained wide popularity: its five parameter structure can fit complex curves dynamics, i.e., curves with multiple inflection points. This is a desirable feature as the BRICS exhibit yield curves with humps which are difficult to approximate with more parsimonious models. Moreover, towards this direction, we also worked on the parameters estimation of the 5F–DRF model, introducing a two–step procedure that significantly improves the overall performance of the method, assuring higher model flexibility and better fitting performances also in presence of market turbulence, which is the core issue where often parametric models fail. With regard to machine learning techniques, on the other hand, we focused on Artificial Neural Networks (ANNs) and, in particular, on FFNNs (Dey, 2016). Thanks to their ability to bring out knowledge from large and not necessarily homogeneous data sets (Di Franco and Santurro, 2020), they have found successful application in a wide variety of fields, including economics and finance (Lopez De Prado, 2018). The rationale for using FFNNs resides in their flexibility and field–proven ability to replicate yield curves dynamics and stylized facts, as testified by recent works of Rosadi et al. (2011), Vela (2013), Posthaus (2019) and Suimon et al. (2020). The FFNNs capability to manage the *ex-ante* uncertainty turns out to be of paramount importance within the BRICS bond market where FFNNs are asked to identify the functional form of the yield curve as well as to overcome the limitations of parametric models in presence of multiple humps. Furthermore, the possibility of customizing FFNNs settings for each country, in order to achieve improvements in their fitting ability for all markets is without any doubt an advantage of using this technique.

In short, our work will try to address the following instances. (i) We try to offer a comprehensive view of the yield curves of the BRICS countries through techniques of consolidate use on developed countries markets. (ii) Applying them on the BRICS, that is a set of five emerging countries, in our opinion, should give proof that the

results are general and not data–dependent. (iii) An in–depth study of the fitting abilities of those techniques on BRICS yield curves, would also help to better address the forecasting issue for future research.

The remainder of the paper is organized as follows. Section 1.2 provides a brief description of both the De Rezende–Ferreira model and the neural architecture employed in our simulation; Section 1.3 contains the results and their discussion; Section 1.4 concludes and offers some suggestions for further research.

1.2 Methodology

1.2.1 The Five Factor De Rezende–Ferreira Model

The De Rezende–Ferreira model assumes to estimate the value RF of the zero–coupon spot rate depending on a 5-parameter vector $\boldsymbol{\beta}$ and a two dimension array $\boldsymbol{\tau}$ to model the set of humps observable on the yield curve dynamics:

$$RF(t, m_k, \boldsymbol{\beta}, \boldsymbol{\tau}) = \beta_0 + \beta_1 \left(\frac{1 - e^{-m_k/\tau_1}}{m_k/\tau_1} \right) + \beta_2 \left(\frac{1 - e^{-m_k/\tau_2}}{m_k/\tau_2} \right) + \\ + \beta_3 \left(\frac{1 - e^{-m_k/\tau_1}}{m_k/\tau_1} - e^{-m_k/\tau_1} \right) + \beta_4 \left(\frac{1 - e^{-m_k/\tau_2}}{m_k/\tau_2} - e^{-m_k/\tau_2} \right) + \epsilon_{m_k},$$

where t is the estimation time, m_k is the maturity over a set of N possible values $m_1, m_2, \ldots, m_N$ that can be both fractions and multiples of the year, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$ and $\boldsymbol{\tau} = (\tau_1 \ \tau_2)'$ are the parameters and decay terms vectors, respectively. In particular, β_0 represents the impact of the long–term component which is constant for every maturity; β_1 and β_2 are the weights associated to the short–term components, β_3 and β_4 are the weights of the medium–term components. Additionally, τ_1 and τ_2 control the convergence speed of the exponential components and determines the maturity at which the medium–term components reach their maximum. By construction, there is a trade-off between the values of τ_1, τ_2 and the effectiveness of the model at long/short term maturities. Big values of τ_j $(j = 1, 2)$, in fact, result in a slow decay and ensure a better fit at long maturities but not in the short term if marked curvatures are present; conversely small values of τ_j $(j = 1, 2)$ get a quick decay and hence a better fit at short maturities, but not in the long run. Finally, ϵ_{m_k} is the error term for all the maturities $m_k, k = 1, \ldots, N$, with:

$$\epsilon_{m_k} \sim \mathcal{N}(0, \sigma_k^2) \ \forall m_k, \quad cov(\epsilon_{m_j}, \epsilon_{m_k}) = 0 \ \ \forall j, k = 1, \ldots, N$$

In our study we used a slight modification of the model discussed in De Rezende and Ferreira (2008). We consider the decay components as time–varying parameters and we run a two–step estimation procedure that at each time t finds the optimal pair $(\hat{\tau}^*_{1t}, \hat{\tau}^*_{2t})$ as the one associated to the lowest Root Mean Square Error (RMSE). This optimal pair is then employed to get $\hat{\boldsymbol{\beta}}^*(t) = (\hat{\beta}^*_{0t}, \hat{\beta}^*_{1t}, \hat{\beta}^*_{2t}, \hat{\beta}^*_{3t}, \hat{\beta}^*_{4t})'$. In contrast to an a priori selection of decay terms, this procedure provides the model with the highest adaptive capability, that is a very desirable feature to use in such a turbulent context like that of the BRICS markets.

For an easier understanding, on following we describe the main steps of the procedure.

1. For each market we define the sets $\Omega_j = \{m_{j,k}\}_{k=1,\ldots,N_j}$ of maturities $m_{j,k}$ with $j = 1,2$ and N_j equal to the sets cardinality. In particular, $m_{1,1}$ is the lower bound of Ω_1 $(m_{1,L})$ and corresponds to the first available maturity of the market, while the upper bound $m_{1,U}$ is, at the same time, the lower bound of Ω_2, that is $m_{1,U} = m_{2,L}$ and it is equal to the straddling maturity between the short and medium–term period. Finally, the upper bound of Ω_2 $(m_{2,U})$ is the longest observed maturity. In our study we set $m_{2,U} = 30$ years, as in general there aren't any bonds traded for longer maturities in the analyzed markets. Values in Ω_1 and Ω_2 ranges between corresponding lower/upper values by proper step sizes Δ_1 and Δ_2. As the step size can affect the overall performance of the procedure, we tried various step sizes in the range [0.25, 0.75] for Δ_1 and [0.25, 1] for Δ_2. After extensive simulations we set $\Delta_1 = 0.75$ and $\Delta_2 = 1$.

2. For each maturity $m_{j,k}$ in the sets Ω_1 and Ω_2 we estimated the parameters $\tau_1(t)$ and $\tau_2(t)$ that maximize the medium term component:

$$[\tau_j(t)(1 - e^{-m_{j,k}/\tau_j(t)})/m_{j,k}] - e^{-m_{j,k}/\tau_j(t)} \qquad k = 1, \ldots, N_j$$

   In this way we get as many curves as the number of maturities.

3. For each time t in the time horizon of length T and for every maturity $m_{j,k}, k = 1, \ldots, N$, keep $\tau_1(t)$ constant and vary $\tau_2(t)$ to estimate by OLS different array sets $\hat{\boldsymbol{\beta}}(t)$; choose then the set $\hat{\boldsymbol{\beta}}^*(t)$ associated to the lowest Sum of Squared Residuals (SSR):

$$SSR(t) = \sum_{k=1}^{N} [y(t, m_k) - \widehat{RF}(t, m_k, \hat{\boldsymbol{\beta}}(t), \boldsymbol{\tau}(t))]^2 \qquad t = 1, \ldots, T$$

with $y(t, m_k)$ and $\widehat{RF}(t, m_k, \hat{\boldsymbol{\beta}}(t), \boldsymbol{\tau}(t))$ being the observed and fitted spot rates respectively.

4. Repeat Step 3 for all $\tau_1(t)$ so that there are as many sets of optimal parameters $\hat{\boldsymbol{\beta}}^*(t)$ as the τ_1 values. Then, select the set with the lowest SSR and fit the yield curve at the desired time t:

$$\hat{\boldsymbol{\tau}}^*(t) = \underset{[\tau_1(t),\ \tau_2(t)]\ \in\ (\Omega_1 \times \Omega_2)}{arg\ min} \left\{ \sum_{k=1}^{N} [y(t, m_k) - \widehat{RF}(t, m_k, \boldsymbol{\tau}(t), \hat{\boldsymbol{\beta}}^*(t))]^2 \right\}$$

5. Repeat Steps 3–4 for each time t $(t = 1, \ldots, T)$, to get the set of T yield curves fitting and the related time series for all the model parameters.

1.2.2 Feed Forward Neural Networks

Feed Forward Neural Networks (FFNNs) are non–linear regression tools which do not require any a priori assumption about the functional form or statistical properties of the data set under examination and can be used to identify and model nonlinearity in the data (Hornik et al., 1989). They are made of information processing units (nodes or neurons) arranged into one or more interconnected layers.

Figure 1.1 shows a graphic representation of an artificial neuron: each input vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$ is processed by a linear combiner (Σ) through the use of a weight vector $\boldsymbol{w} = (w_1, \ldots, w_n)'$, and hence transformed with the aid of a proper activation function (usually the sigmoid). At this stage a bias value can be inserted to delay the triggering of the activation function.

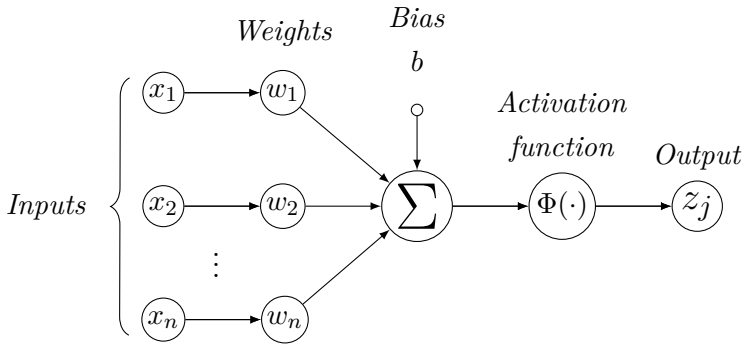


Figure 1.1 *– Representation of a node with n input features*

Nodes are grouped in three types of layers: the Input layer with nodes supplying the input features to next layers consisting of one or more Hidden layers which, in

contrast to Input/Output layers, are not in direct contact with either the network input or the output. The response value (z_j) is sent either to another hidden layer (if any) or to the Output layer that is the neurons layer which generates the final response of the network.

In order to obtain the desired output, the network must undergo a learning process which identifies the optimal weights configuration (Wilamowski and Irwin, 2011): increasing or decreasing the weights values by means of proper learning algorithms changes the strength of nodes connections and directly affects the network capability to learn the input space features. The FFNN makes use of a supervised learning algorithm: information and correct target results are available and presented to the network which tries to define the optimal weights configuration, so that the network response is as close as possible to the correct output. In detail, we used the Backpropagation Algorithm—BPA (Rumelhart et al., 1986). This is an iterative procedure through a certain number of cycles (epochs), each including two phases: a forward stage with the network, initialized at random, generating the output signals (responses), and a backward phase. In the latter the network responses are compared to the target values, the error is back–propagated from the output layer through the hidden ones towards the inputs, and then used to update the networks coefficients to reduce the error at the end of the next forward phase. The process goes on until the optimal combination of weights that minimizes a loss function (e.g., the Mean Squared Error—MSE) is determined according to the gradient descent criterion, so that the estimated output values are the closest to the target output.

1.3 Empirical Analysis

1.3.1 Data

Our dataset consists of daily returns for government zero–coupon bonds (ZCB) of the BRICS countries, as summarized in Table 1.1, where for each country we provided the Start and the End of the observation period, the overall number of observations and the dataset source, that is either TRD (Thomson Reuters Datastream) or CBR (The Central Bank of the Russian Federation).

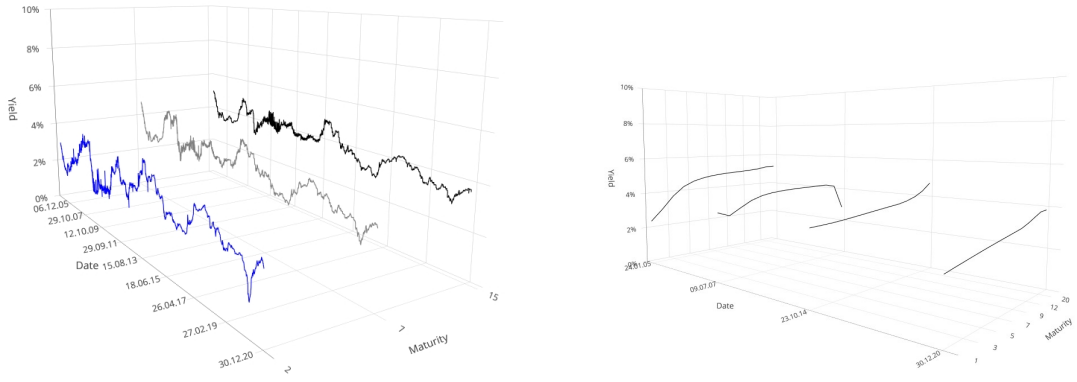
As previously said, different maturities could be used to represent the yield curve, depending on the liquidity conditions and availability of information of the analyzed market: for example Diebold and Li (2006) examined maturities from 3 to 120 months (US), while De Rezende and Ferreira (2008) focused on maturities in the range 1 to

60 months (Brazil), and Caldeira et al. (2010) from 1 to 33 months (Brazil). In this work we considered maturities from 3 months (i.e., 0.25 of the year) to 30 years.

Table 1.1 *– Dataset description*

Country	Period		Nº of Observations	Source
	Start	End		
Brazil	30/09/2011	30/12/2020	2128	TRD
Russia	04/01/2003	30/12/2020	4578	CBR
India	14/02/2012	30/12/2020	2185	TRD
China	24/01/2005	30/12/2020	3818	TRD
South Africa	18/02/2011	30/12/2020	2472	TRD

Combining the available maturities with the observed time we obtained a tensor whose number of rows is equal to the number of analyzed days and with the number of columns equal to the number of maturities. In this way, for each maturity it is possible to observe the evolution of the spot rates time-series: Figure 1.2a provides an example on the Chinese market with the maturity set at 2, 7 and 15 years respectively; moreover for each day it is possible to extract the yield curve varying the maturities: in Figure 1.2b we give an example on the Chinese market with t set to 24/01/2005, 09/07/2007, 23/10/2014 and 30/12/2020.



(a) ***(b)***

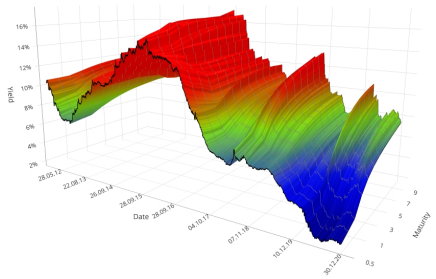
Figure 1.2 *– Behaviour of the daily rates for maturities 2, 7 and 15 years for China in the period 24/01/2005 – 30/12/2020 (a), and yield curve shapes for China (b) observed in t=24.01.2005; 09.07.2007; 23.10.2014 and 30.12.2020, respectively.*

A closer look to Figure 1.2a suggests the presence of spikes in the yield curve at various maturities. Such variability is confirmed if we turn to Figure 1.2b, which offers a snapshot inside how many different behaviours (flat, normal, inverted) the yield curve can show depending on the time assumed as observation point.

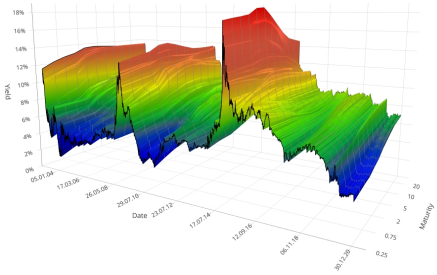
Furthermore, varying both t and m for each country it is also possible to build a 3D surface chart as shown in Figure 1.3 for all the BRICS, where the x-axis reports the time t, the y-axis shows the maturity m expressed in fractions (multiples) of the year and the z-axis the observed level of the interest rate at each time t and for each maturity m.

Looking at the shapes of the yield surfaces in Figure 1.3 suggest that they are the result of quite different instances, that is they have been affected not only by external conditions such as the global crisis or the drop in the commodities demand, but also by internal drivers, such as policy decisions, unemployment and recession in general. In the case of Brazil, (Figure 1.3a), for instance, pronounced spikes are observable since 2015, when the country began to be interested by a crippling two–year recession (in 2015 and 2016) which was only partially recovered in next years. The booms and busts of the yield surface after 2017 bear witness of the uncertainty dominating the market in those years. In the case of Russia (Figure 1.3b), on the other hand, the highest turbulence in the yield surface corresponds to the period between 2014 and 2016, when the Russian economy suffered from a currency crisis caused by the collapse of oil prices and the country's engagement in the conflict in Ukraine. In the case of India (Figure 1.3c), instead, the mostly flat surface is due to the persistence of economic regression during the whole 2010s. Notably, we can observe an inverted shape of the yield surface corresponding to the period 2012–2014, when India underwent the worst slowdown of the decade in the manufacturing and mining sectors, both of which labour intensive crucial sectors for the growth of other sectors. Conversely, at the end of 2020 the surface turned to a normal behaviour, with lowest (higher) interest rates associated to lower (higher) maturities; how much this stage is solid will depend on the country capability in replying to ongoing inflation of fuel, food prices as well as rising urban unemployment. Similar considerations can be extended to the discussion of the yield surface of South Africa in Figure 1.3e: in the period 2011–2015, the surface is extremely flat, in connection to higher turbulence in the markets. In this period, in fact, like all economies dependent on commodity prices, the South African economy has been exposed to two highly correlated external shocks: the slowdown in Chinese demand for commodities, which has become the leading destination for South African exports in recent years and the ensuing decline in iron ore. Finally, if we turn the attention to

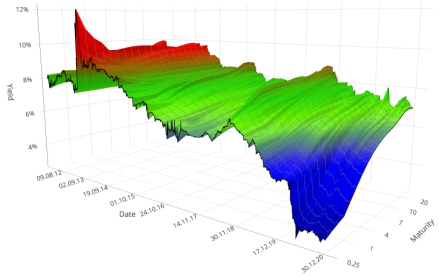
Figure 1.3d, we can observe that highest variability corresponds to the period from June 2007 to September 2010, when the Chinese financial market faced the Global Financial Crisis, while the flattening of the surface in the following years is due partly to the trade war with the USA and partly to central government driving the country's transition towards an economy led by consumption and services, rather than one driven by exports and investment.



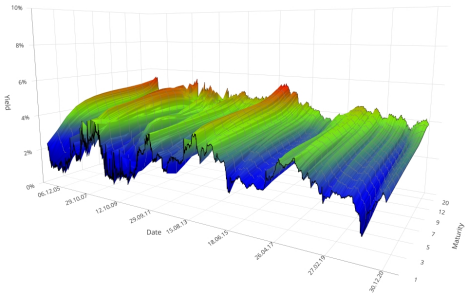
(a) Brazil



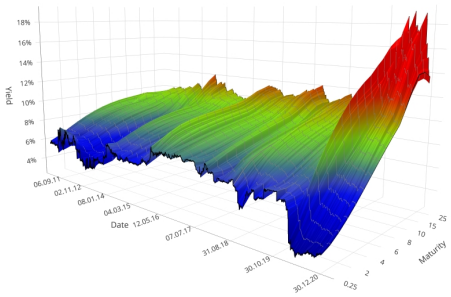
(b) Russia



(c) India



(d) China



(e) South Africa

***Figure 1.3** – Zero-Coupon yield surfaces of the BRICS countries in the monitored period*

Overall, we can preliminary conclude that BRICS countries offer the ground for testing the effectiveness of fitting methods in presence of sensitively varying conditions.

For each country data we then run simulations with the 5F–DRF according to the guidelines given in Section 1.2.1, while for what is concerning the FFNN, the final output was obtained partitioning the countries market data into training (70%), validation (15%) and testing (15%) sets. We defined the number of input, hidden and output neurons depending on the analyzed country; in particular, the quantity of I/O nodes corresponds to the number of available maturities. Regarding the number of intermediate layers and nodes, as there is no a precise rule to select their best combination and the choice is data–dependent (Lantz, 2019), we followed a trial and error approach analyzing different configurations (i.e., with one or more hidden layers). The best performance in terms of determination coefficient R^2 was obtained with the network architectures summarized in Table 1.2 for each BRICS country.

Table 1.2 *– Number of layers and neurons for each BRICS country assuring the best fitting performances.*

Country	Hidden Layer	Input/Output Nodes	Hidden Nodes
Brazil	1	12	9
Russia	1	12	13
India	1	18	10
China	1	14	11
South Africa	1	16	12

All the FFNNs were trained with the backpropagation learning rule; the learning process was cut after 1000 epochs, i.e., after presenting each training set to the network 1000 times. The fitting accuracy of both models was evaluated comparing the Mean Square Error (MSE) and Root Mean Square Error (RMSE) of both the 5F–DRF model and the FFNN.

1.3.2 Comparison of the 5F-DRF and FFNN Models Fitting Performances

In this section we present the results of the empirical estimation of the term structure. The analysis was carried out using R 4.0.4 and a freshly-new R package (Castello and Resta, 2019) with estimation routines implementing the 5F-DRF model, while the MATLAB R2021a (9.10.0) Neural Network Toolbox was employed to run the FFNNs. All the code is available for download in the Zenodo repository (https://zenodo.org/record/5814658 (accessed on January 3, 2022)).

Table 1.3 shows the average estimated parameters values for the time–varying 5F–DRF.

Table 1.3 *– Average estimated parameters values for the time–varying 5F–DRF model*

	Brazil	**Russia**	**India**	**China**	**South Africa**
β_0	11.8010	8.7710	6.0058	2.5294	9.4784
β_1	200.3699	-0.2674	-4084.0492	662.6816	15.8301
β_2	-203.2936	-2.5483	4085.0426	-663.1078	-19.2812
β_3	4.3508	0.5273	-86.0267	31.5497	10.0752
β_4	4.8412	0.3619	-81.2562	37.91.29	20.9306
τ_1	0.8246	1.079	1.9792	2.5018	2.4239
τ_2	3.5799	6.1704	10.376	11.4936	9.2533

The estimated values emphasise the different role played in the yield curve of BRICS by short/medium and long term components. The value of Brazilian β_0, which addresses the long–term effect, is greater than those of Russia and South Africa by 35% and 24%, and by 96% and 366% than those of India and China, respectively. Clearly, these values reflect the different perception of long–term expectation in the observed countries. Again, if we turn to parameters associated to short term components of the yield curve (β_1 and β_2), we observe that they differ significantly from one country to another: the scale of values is in the (negative) tenths for Russia, in the dozens for South Africa, in the hundreds for Brazil and China with alternate signs, in the thousands for India. Although so different in scale, values for Brazil, India and China when compared to corresponding β_0 , β_3 and β_4 highlight the importance of short–term expectation in the yield curve of those countries. This aspect is more shaded in the case of South Africa and ever more so for Russia. Mid–term parameters (β_3 and β_4) are in turn different in scales in various BRICS countries, with higher values associated to India and China, and lower values corresponding to Russia. Overall the estimated values point out how, despite BRICS are often referred as a compact group of countries, the underlying financial and economic drivers substantially differ and they are reflected in the behaviour of the yield curve.

Moving to the results comparison, we obtained a very accurate fit with both the models, as it can be seen in Figure 1.4, showing the average yield curve against the average fitted ones generated by both the 5F–DRF and the FFNN.

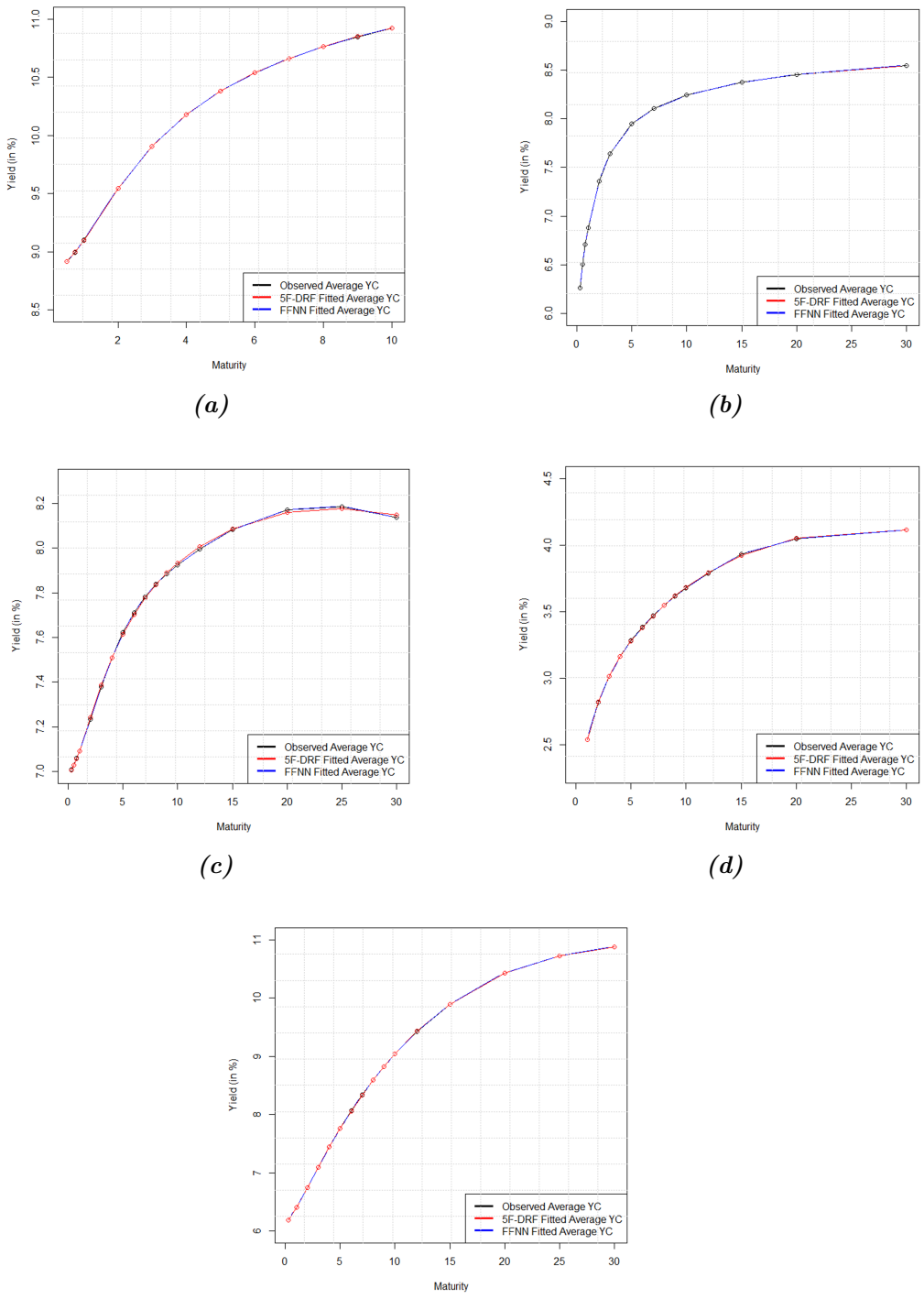
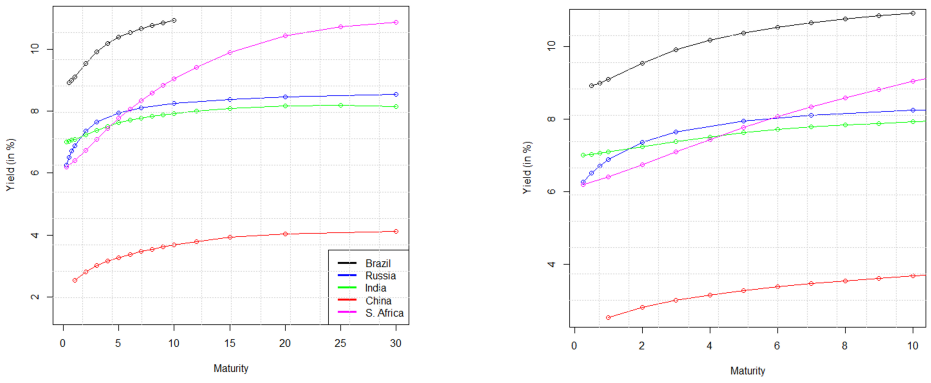


Figure 1.4 – *Average observable yield curve against average yield curves generated by the 5F-DRF (red) and FFNN (blue) models in the BRICS market.*

At first sight it seems not possible to distinguish the average observable curve from the fit provided by both the 5F–DRF and the FFNN. Besides, a common feature for the two methods seems the capability to fit curves with quite different shapes as well as in different variation ranges. Indeed, despite the fact that all the curves appear to share similar shapes, that is increasing as function of the maturity, the slope of the average yield curve of South Africa in Figure 1.4e grows faster than in other cases. This curve, in fact, varies in a range of five percentage points while the range of variation of the other yield curve is much smaller. Conversely, the average yield curve of India, in Figure 1.4c is the flattest one with a range of variation of only 1.2 percentage points. Moreover, both the methods were able to capture different variation ranges: in the interval $[9, 11]$ for Brazil, $[6, 9]$ for Russia, $[7, 8.2]$ for India, $[2.5, 4]$ for China and $[6, 11]$ for South Africa. Our first conclusion is that the analysis of the average yield curves does not provide sufficient evidence to uncover which is the best fitting method. Indeed, the *lesson learned* from this first exploration into the results is that both methods work very well in keeping very different behaviors. In Figure 1.5, in fact, we offer a direct comparison on the average observable yield curve of the BRICS countries where it is possible to look at the difference in both the variation range and the shape of the curves, as we have already highlighted in the above rows. This feature makes the fitting capabilities of the two methods even more valuable because they were able to capture the dynamics of all the curves, despite the difference in both the maturity spectrum and the steepness of the curve.



(a) ***(b)***

Figure 1.5 *– Plot of the BRICS average observable yield curves (a) and a zoomed-in area (b) which covers the maturity spectrum common to all the examined countries.*

In search of more clues, we then turned to analyze the average residuals that are shown in Figure 1.6.

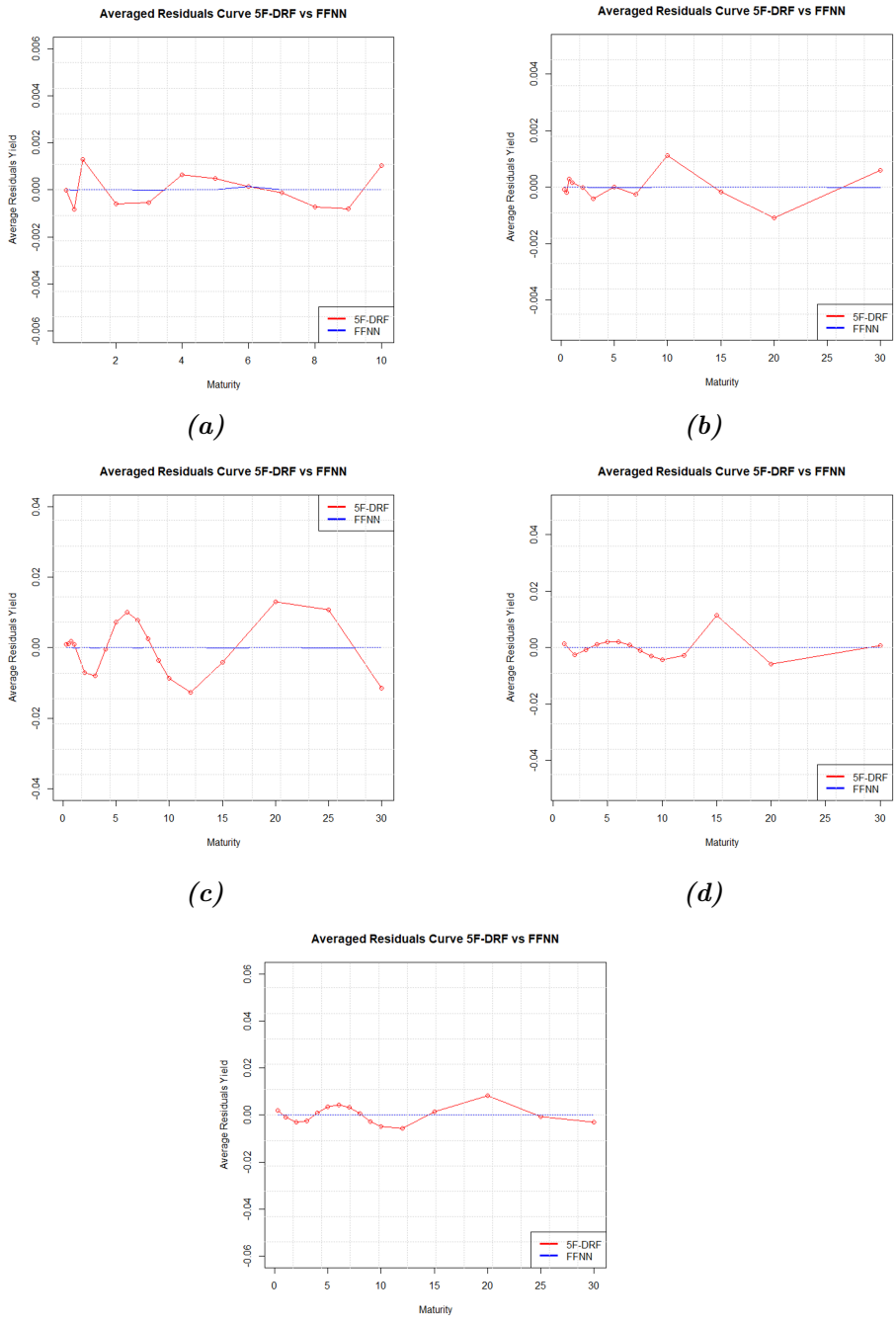


Figure 1.6 *– Comparison of average residuals curves generated by 5F–DRF and FFNN.*

The average residuals are very close to 0 for all the BRICS countries: the range of variation is between $\pm 2 \times 10^{-3}$ for Brazil and Russia while for India, China and South Africa it lies between $\pm 2 \times 10^{-2}$. These results highlight how both models are

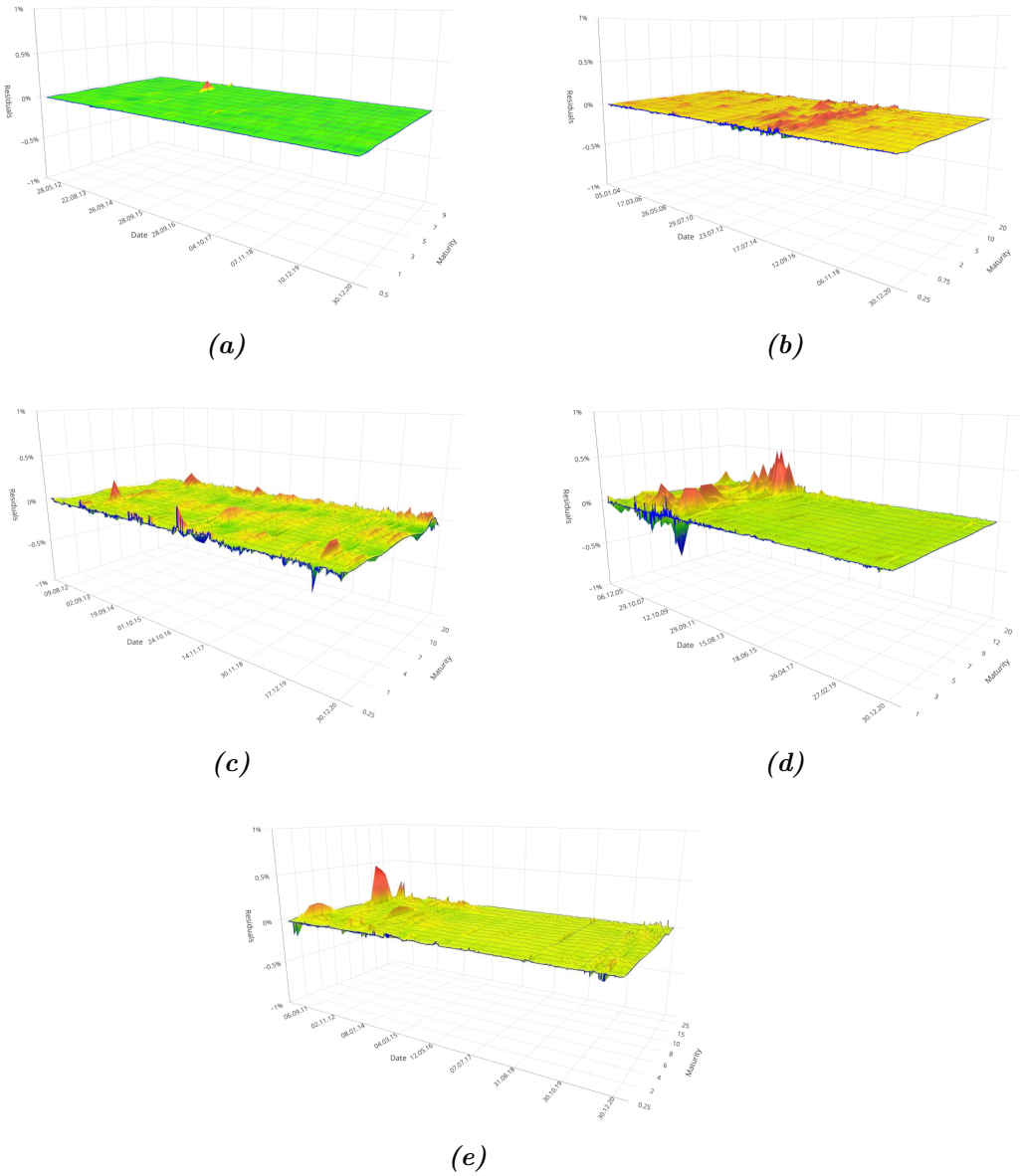
characterized by excellent average fitting capabilities. However FFNNs residuals are lower at least by a factor of 100 compared to those of the 5F–DRF model which are characterized by higher oscillations. Taking for example the case of China at the maturity 1 and 30, for $m = 1$, the errors are equal to 1.3×10^{-3} and -3.7×10^{-7} for the 5F–DRF and FFNNs respectively, and equal to 7.7×10^{-4} and -1.2×10^{-6} for $m = 30$.

Table 1.4 *– Main statistics for MSE and RMSE associated to 5F-DRF and FFNN models applied to the BRICS countries.*

		MSE		**RMSE**	
		5F-DRF	**FFNN**	**5F-DRF**	**FFNN**
Brazil	Mean	1.0511×10^{-5}	7.6401×10^{-7}	2.5536×10^{-3}	6.9986×10^{-4}
	SD	6.3678×10^{-5}	2.6374×10^{-6}	1.9980×10^{-3}	5.2378×10^{-4}
	Min	1.4523×10^{-8}	7.6170×10^{-9}	1.2051×10^{-4}	8.7275×10^{-5}
	Max	2.8513×10^{-3}	1.0155×10^{-4}	5.3398×10^{-2}	1.0077×10^{-2}
Russia	Mean	6.8143×10^{-5}	2.3630×10^{-8}	6.1318×10^{-3}	1.2773×10^{-4}
	SD	1.5740×10^{-4}	5.8504×10^{-8}	5.5273×10^{-3}	8.5538×10^{-5}
	Min	4.5277×10^{-7}	6.5816×10^{-10}	6.7288×10^{-4}	2.5655×10^{-5}
	Max	1.8731×10^{-3}	1.8621×10^{-6}	4.3279×10^{-2}	1.3646×10^{-3}
India	Mean	3.3058×10^{-4}	1.0865×10^{-6}	1.3763×10^{-2}	2.6338×10^{-4}
	SD	5.6081×10^{-4}	4.8029×10^{-5}	1.0789×10^{-2}	1.0087×10^{-3}
	Min	5.3692×10^{-7}	7.1060×10^{-10}	7.3275×10^{-4}	2.6657×10^{-5}
	Max	8.6687×10^{-3}	2.1984×10^{-3}	9.3106×10^{-2}	4.6887×10^{-2}
China	Mean	7.1533×10^{-4}	6.7902×10^{-8}	1.4338×10^{-2}	1.7081×10^{-4}
	SD	3.0070×10^{-3}	4.1125×10^{-7}	2.2551×10^{-2}	1.9681×10^{-4}
	Min	4.5227×10^{-8}	2.0021×10^{-10}	2.1267×10^{-4}	1.4149×10^{-5}
	Max	6.3573×10^{-2}	1.6570×10^{-5}	2.5214×10^{-1}	4.0706×10^{-3}
S. Africa	Mean	3.3058×10^{-4}	1.0865×10^{-6}	1.3763×10^{-2}	2.6338×10^{-4}
	SD	5.6081×10^{-4}	4.8029×10^{-5}	1.0789×10^{-2}	1.0087×10^{-3}
	Min	5.3692×10^{-7}	7.1060×10^{-10}	7.3275×10^{-4}	2.6657×10^{-5}
	Max	8.6687×10^{-3}	2.1984×10^{-3}	9.3106×10^{-2}	4.6887×10^{-2}

To evaluate and compare the models fitting abilities we also calculated the MSE and RMSE for the observed yield curves for each maturity, reporting main statistics in Table 1.4.

The results on the one hand show that both models are characterized by very low values of the MSE. This confirms that on average they both are able to fit the wide variety of shapes exhibited by the yield curves. On the other hand, the results clearly underline the superior fitting abilities of FFNNs on the 5F–DRF model.



(a) ***(b)*** ***(c)*** ***(d)*** ***(e)***

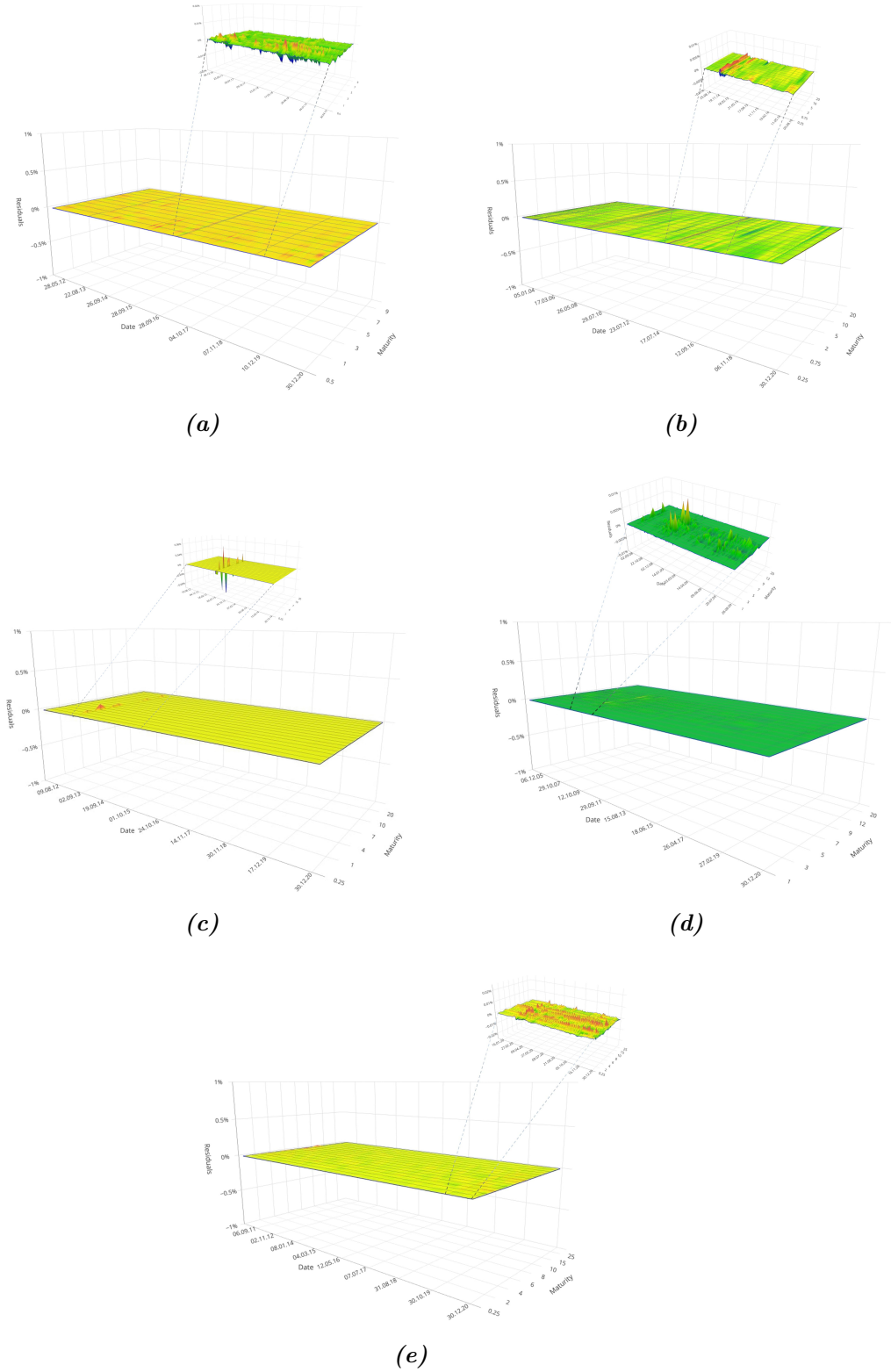
Figure 1.7 – *Surface of daily residuals for the yield curve obtained with the 5F–DRF in the BRICS bond market of Brazil (**a**), Russia (**b**), India (**c**), China (**d**), South Africa (**e**).*

Further evidence of the higher fitting abilities of FFNNs is given in the 3D plots shown in Figures 1.7 and 1.8 which represent the residuals surface generated by the 5F–DRF and FFNN models respectively, obtained varying both the time t (x–axis) and the maturity m (y–axis). For the FFNN the error surface also includes a zoomed-in area to highlight the residuals magnitude otherwise unobservable at the same scale used to monitor the errors behaviour in the 5F–DRF.

In principle, we can conclude that both models provided very good fitting performances. The residuals did not exhibit systematic behaviour, i.e., neither incorrect zeroing or cyclical behaviour caused by wrongly specified models; they took absolute values of very small magnitude with fairly rare spikes mainly concentrated on periods characterized by greater market volatility. In this regard, the error surface of both models contains additional clues leading us to the following conclusions:

- (a) FFNNs perform better than the 5F–DRF model for in sample fitting of the yield curve of the BRICS countries.
- (b) The reason of (a) is in a better adaptability of the FFNN to both internal and external shocks.

The surface generated by the parametric model (5F–DRF), in fact, highlights that more pronounced residuals values are associated to higher market turbulence, as discussed analyzing the yield curve surfaces in Figure 1.3. This is, for instance, the case of Russia, with maximum values in the error surface of the 5F–DRF corresponding to the period between 2014 and 2016. Again, if we turn the attention to Figure 1.7d, we can observe that maxima in the error surface correspond to the period from June 2007 to September 2010, when the China was struggling with the Global Financial Crisis. In the case of India (see Figure 1.7c), instead, we can observe that the whole error surface is disseminated by spikes, mainly concentrated at lower maturities. This clearly indicate an underestimation (overestimation) of the observable values; as a matter of fact, the behaviour of the Indian yield curve at various maturities was a bit trickier to fit than for other countries in the sample: the economic regression probably affected the volatility of financial markets. Similar remarks hold also for South Africa and its error surface in Figure 1.7e: error peaks raised at the maximum level once again in connection to higher turbulence in the markets. The lone voice in this roundup comes from Brazil, with a flat error surface for the 5F–DRF as it can be seen in Figure 1.7a, and more pronounced spikes (however not exceeding the range ± 0.001) in the zoomed–in area of Figure 1.8a, during the period 2015–2018.



(a) ***(b)*** ***(c)*** ***(d)*** ***(e)***

Figure 1.8 – *Surface of daily residuals for the yield curve obtained with the FFNN in the BRICS bond market. As the error surface is very flat, the inset shows a zoomed–in area highlighting the error fluctuations otherwise not visible at the same scale employed to visualize the error surface of the 5F–DRF. From top to bottom and from left to right,* (***a***) *is associated to Brazil,* (***b***) *to Russia,* (***c***) *to India,* (***d***) *to China,* (***e***) *to South Africa.*

In general, we can conclude that 5F–DRF model suffers for underestimation (overestimation) issue when used to interpolate curves with pronounced oscillations at certain maturities. These are common for parametric model, and even the use of a more efficient parameters estimation can only reduce the error but not at the levels provided by non–linear black box technique like the FFNN. At the same time, in fact, FFNNs provided evidence of being a more flexible tool for in–sample fit of the term structure of interest rates: the residuals generated by the FFNNs are considerably lower, than those of the 5F–DRF and less influenced by the market turbulence in every analyzed situation. They therefore denote a greater ability to replicate almost all the patterns exhibited by the term structure curve, even in cases of strong market fluctuations or downturn.

1.4 Conclusion

Motivated by the important role played by the term structure of interest rates we investigated and compared the in–sample fitting abilities of two distinct methods, the Five Factor De Rezende–Ferreira (5F–DRF) model and Feed–Forward Neural Networks (FFNNs) when applied to the data of BRICS countries. For the parametric approach we discussed the use of time–varying decay terms to ensure more flexible parameters and hence get higher interpolating performances. Focusing on in–sample rather than out of sample was not limited, in our opinion, as bond prices reflect market participants' views on interest rate levels in a forward–looking way.

The results may be analyzed under two reading keys. First, we offered a comprehensive study of the BRICS yield curves all at once. In this respect, we outlined how despite BRICS countries are often viewed as a compact set, this is not true, as reflected by the different behaviour of related yield curves. This aspect was accurately kept by the 5F–DRF model, as it can be seen by looking at the average values of estimated parameters: the values, in fact, highlight the different weight associated to short/mid and long–term components of the yield curve. Second, we highlighted very high in–sample fitting capabilities of both the models for all the examined countries. Nevertheless although the 5F–DRF is the most flexible model in the Nelson–Siegel Family and these features have been further enhanced by way of time–varying parameters, the empirical evidence has clearly shown the superior capability of FFNN in interpolating the behaviour of the yield curve; this was also confirmed by comparing the average behaviour of monitored and interpolated yield curves as well as examining the average residuals and the residual surface of both the 5F–DRF and the FFNN. In

fact, the FFNN perfectly adapted to all the typical yield curve shapes, even to the most twisted ones with multiple inflection points like in the Indian and Russian case. Moreover FFNNs efficiently replicated all the features of the examined term structures, being flexible enough to overcome the common limitations of parametric models in presence of booms and busts. The greater ability of the FFNN was confirmed by the MSE and RMSE associated to the fit, with values ranging, in the worst case, between ± 0.02 for the 5F–DRF and between ± 0.0002 for the FFNN. Another advantage of FFNNs was related to the estimation process which is less time–consuming than for the 5F–DRF, because it requires a lower amount of parameters during the network calibration process. The possibility of customizing FFNNs settings for each country, in order to achieve improvements in their fitting ability for all markets was without any doubt an additional advantage of using this technique.

Based on the arguments set out above, we therefore conclude that the FFNN is a better and flexible tool for the in–sample fit of the yield curve in all the BRICS markets. Nevertheless we do not underestimate some limitations of our approach and mainly the fact that we performed in–sample fitting. To such aim, future plans include the extension of our conclusions by comparing models performance out–of–sample and a deepest investigation of the potential of our procedure to estimate time–varying parameters in the 5F–DRF compared to alternative solutions discussed in the more recent literature on parametric models.

Chapter 2

Optimal time varying parameters in yield curve modeling and forecasting: A simulation study on BRICS countries

2.1 Introduction

As widely known, the term structure of interest rates depicts the relationship between different times to maturity and the market remuneration rate (interest rate). Its graphical companion is the yield curve, that plots the interest rates of bonds with equal credit quality at different maturities: its shape and time changes are conventionally considered a key indicator for the economic outlook of a country (Chadha et al., 2014). The yield curve can be used to represent either spot rates, that is the yield associated to a zero-coupon bond from now to maturity, or forward rates, that is the yield of a zero–coupon bond between two future dates. Indeed, the yield curve plays a valuable role as alerting tool for inflation, possible recession or upturn of the economy (Gürkaynak and Wright, 2012) and it can show several shapes depending on market volatility; due to this pivotal role, various models have been suggested to analyze, model and predict it.

Actually the most popular models are those in the so–called Nelson–Siegel (NS) family, pioneered by Nelson and Siegel (1987) and since then by Bliss (1996) who discussed an extension with an additional decay parameter, Svensson (1994) with a four–factor model including a further curvature term and De Rezende and Ferreira (2008) introducing a five–factor model with two slope terms instead then only one.

A dynamic version of the NS model was then suggested by Diebold and Li (2006) who considered parameters as time–varying latent factors to achieve a more effective forecasting of the yield curve. Further attempts to increase the flexibility of NS models are described in Koopman et al. (2007), who examined time–varying factors loading and volatility, Christensen et al. (2007, 2009) who introduced the new class of Affine Arbitrage–free Dynamic Nelson–Siegel models and in Ullah (2017) who discussed time–varying asymmetric volatility. The fitting and forecasting abilities of NS models have been also studied in De Pooter et al. (2010) and Fernandes and Vieira (2019) whose models also include the interaction between the yield curve and the economic system.

For what it concerns empirical studies, there is a plenty of works dealing with the use of parametric models both for in–sample fitting and out–of–sample forecasting of the yield curve: Linton et al. (2001) analyzed the U.S. bond market, Chou et al. (2009) compared the modeling performances of different parametric models for Taiwan Government Bonds, Hoffmaister et al. (2010) examined a dynamic parametric representation of the Central and Eastern European Countries yield curves, Kang (2012) forecasted the term structure of Korean Government bond yields with various types of dynamic parametric models, Lorenčič (2016) and Nagy (2020) analyzed the estimation abilities of parametric models on the Austrian and the Hungarian term structures with missing data, respectively. Finally, Luo et al. (2021) fitted and predicted U.S. Treasury yield curves using the Dynamic Nelson–Siegel Model with random level shift parameters, while Umar et al. (2022) applied NS models to the countries in the Group of Seven to investigate the risk transmission mechanism. The overwhelming majority of the above studies focused on economies with enhanced resilience to economic downturns and relatively stable term structure dynamics, i.e. absence of spikes or drops in the level of interest rates. In this respect it makes sense to use models based on constant decay–terms and thus sub–optimal parameters for yield curve estimation and forecasting. However, such an approach may lead to inconsistent results when applied to structures that show a volatile behavior with frequent trend inversions, jumps and/or falls, like those observed in emerging markets.

Our work nests in the above debate and investigates the use of parametric models characterized by optimal time–varying decay factors and parameters for in–sample fitting and out–of–sample forecasting within emerging markets. The rationale arises because various studies highlighted that those markets are more sensitive to both endogenous and exogenous shocks (Chiţu et al., 2018, Bhattarai et al., 2021). In our opinion, the use of optimal decay terms and parameters endows these models with necessary flexibility to manage the challenging dynamics characterizing these markets.

Existing contributions related to emerging economies so far were oriented either to model (e.g. Zoricic and Orsag, 2013, Petousis and Barr, 2016, Chouikh et al., 2017, Lartey and Li, 2018, Lartey et al., 2019, Ertan et al., 2020) or to predict (e.g. Caldeira et al., 2016, Poghosyan and Poghosyan, 2019) the term structure of single countries. However, this approach has an evident drawback since it is not necessary true that the methodology which is appropriate for a country is still good for other ones.

In the light of the above, the scope of our research is twofold: on the one hand we discuss the use of optimal factors and parameters for models in the Nelson–Siegel Family and test their capabilities outside the *comfort zone* of developed and stable markets; while on the other we provide a comprehensive study focused on in–sample modeling at first, and then on out–of–sample predictions in order to asses the overall effectiveness of the proposed methodology.

In this respect, our paper contributes in several ways. First, within the dynamic framework discussed in Diebold and Li (2006), we use both the Three Factor Dynamic Nelson–Siegel (3F–DNS) and the Five Factor Dynamic De Rezende–Ferreira (5F–DRF) models and we discuss an estimation technique based on time–varying decay factors that leads to significant enhancements of both models fitting abilities. Furthermore we focus on the forecasting issue following an alternative approach to direct spot rates prediction: we run day–ahead predictions of the models parameters and we then derive the corresponding yield curve using various forecasting methods: the Univariate Autoregressive process AR(1), Trigonometric Seasonal Box–Cox Transformation with ARMA residuals Trend and Seasonal Components (TBATS) and the Autoregressive Integrated Moving Average (ARIMA) that we combined to a Nonlinear Autoregressive Neural Network (NAR–NN). To the best of our knowledge, this study is the first to test the potentials of TBATS and ARIMA–NARNN models with the purpose of forecasting the term structure of interest rates. Additionally, we manage those methods as Local Data Generating Processes, that is we calibrate the models using a quite limited range of values which are next to the prediction period. In this way we give greater emphasis to the information content of the period close to that of forecast. This approach allows to considerably reduce the volatility of predictions which may result from the use of considerable amounts of data given the characteristics of emerging markets. Finally, we conduct our analysis on a pool of 5 countries, that is Brazil, Russia, India, China and South Africa, considering a wide time span of at least 10 years of daily observations thus including global schocks and the most recent financial crises. This offers a breeding ground for the stress–testing of the proposed framework.

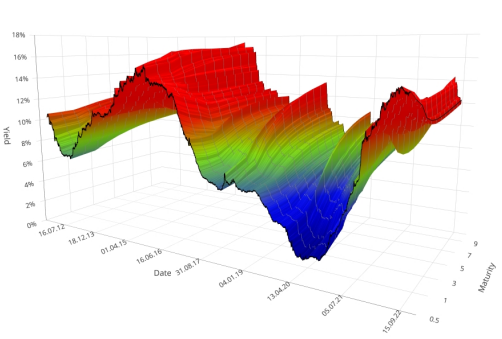
The remainder of the paper is organized as follows. Section 2.2 describes materials and models and its divided into three parts: a section presenting the data employed in the simulation; a part introducing the 3F–DNS and 5F–DRF models and after that the approach followed to estimate each models parameters. In Section 2.3 we then discuss the fitting results, while Section 2.4 discusses the forecasting of the BRICS yield curves. Section 2.5 closes the paper with final remarks and outlooks to address further research.

2.2 Materials and Models

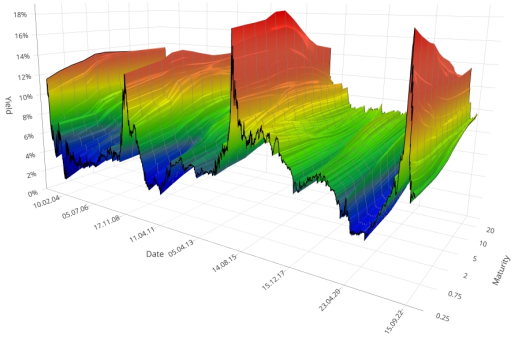
2.2.1 Data

The data set in use consists of daily spot rates for the government zero–coupon bonds (ZCB) of Brazil, Russia, India, China and South Africa, i.e. the so–called BRICS. We examined maturities in the range from 3 months (i.e. 0.25 of the year) to 30 years. The data were collected from Thomson Reuters Datastream (TRD) and the Central Bank of the Russian Federation (CBR). The observation period, is not homogeneous for the examined countries; starting points are 09/2011 for Brazil, 01/2003 for Russia, 02/2012 for India, 01/2005 for China and 02/2011 for South Africa; on the contrary, the ending period, 09/2022, is common to all markets. As a consequence, the sample period ranges between a minimum of 10 and a maximum of 19 years, depending on the examined market, for an overall amount of 2557 observations for Brazil, 4995 for Russia, 2631 for India, 4234 for China and 2908 for South Africa.

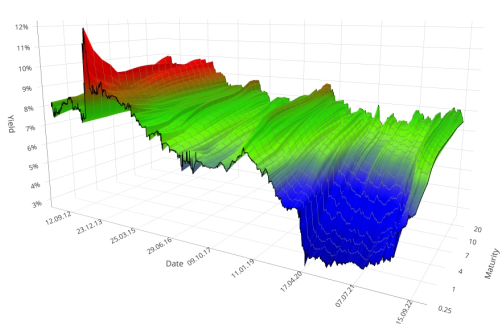
The time span was chosen to include critical situations such as global shocks and events like the Subprime Mortgage crisis of 2007–2009 and the consequent Great Recession, the 2015–2016 Chinese stock market crisis as well as the oil and pandemic turmoil of 2020 and the more recent geopolitical crisis of 2022 induced by the conflict in Ukraine. Those events had a significant impact on BRICS securities market, as resulting from the behavior of the yield curve, provided in Figure 2.1 where we plot the term structures 3D surface for each market obtained varying both the time (on x–axis) and the maturity (on the y–axis). The Euro Zone (EU) term structure surface is also shown for benchmarking purposes. Given the same historical time frame and market events, it can be noted that the EU surface is flatter and less volatile.



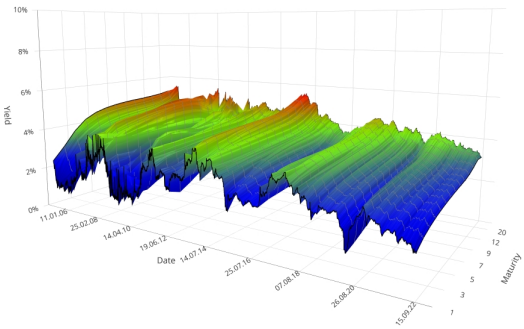
(a) Brazil



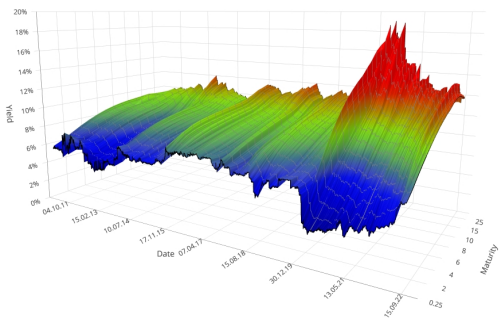
(b) Russia



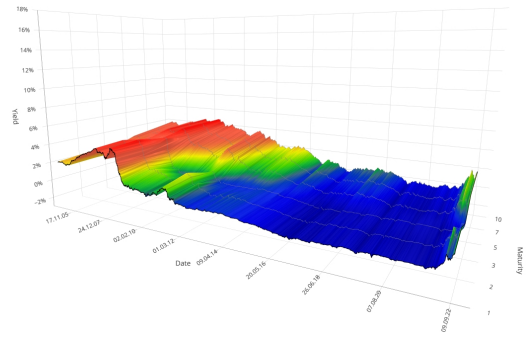
(c) India



(d) China



(e) South Africa



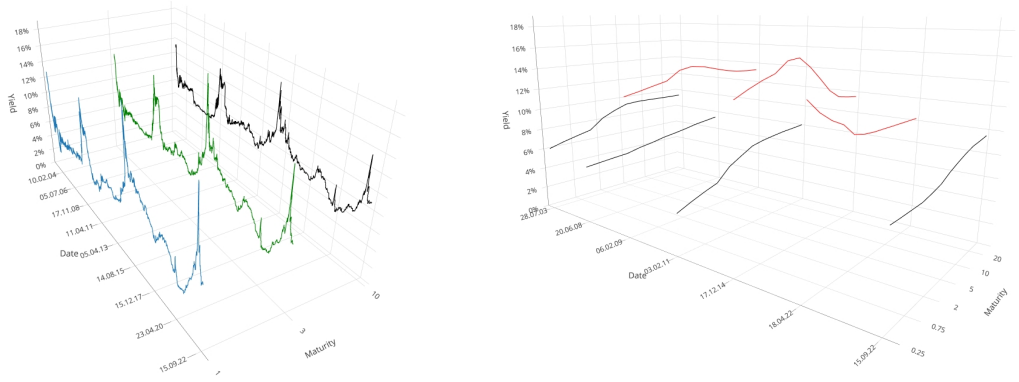
(f) Euro Zone

Figure 2.1 *– From top to bottom in clockwise sense, the yield curve surface for Brazil (a), Russia (b), India (c), China (d), South Africa (e) and Euro Zone (f). Time is represented on the x–axis, while the tenor (expressed in fractions or multiples of the year) is on the y–axis and the yield on the z–axis.*

Furthermore, the 3D surface plot makes possible to highlight all the yield curve shapes occurring in the period that is almost all typical patterns: the normal trend, upward sloping and concave, indicating a quite stable economic outlook; the flat

behavior, with short–term rates similar to medium and long–term ones, indicating a possible slowdown of the economic system; the inverted curve, with short–term rates higher than long–term ones, usually interpreted as a signal for recession; and the S-shaped curve characterized by sudden and marked multiple changes in the slope and curvature.

Examining, for instance, the Russian bond market shown in Figure 2.1(b), it is possible to observe the absence of structural changes and shocks from mid–2003 up to early 2008, as well as from 2010 to 2014 and, more recently, from 2016 till 2020, with average yield values at different maturities in the range 4.88% - 8.44% and curve shapes mostly upward or flat. However, these conditions are broken by unstable periods characterized by greater volatility and the presence of relevant jumps for all maturities, as it can be also seen in Figure 2.2 *(a)*.



(a) ***(b)***

Figure 2.2 *– The left panel (a): Plot of the Russian daily rates setting the maturity to 1, 3, and 10 years. The right panel (b): Yield curve shapes extracted from the Russian term structure 3D surface in (b). Time is represented on the x–axis, while the tenor (expressed in fractions or multiples of the year) is on the y–axis and the yield on the z–axis.*

The plot shows three *slices* of the yield surface of Figure 2.1(b), i.e. three time series extracted at the maturities 1, 3 and 10 years. In all the three cases it is possible to highlight instability patterns. The first one occurred in the period June 2008 – May 2010, a kind of follow–up of the recession due to the Subprime crisis, with a growth of approximately 11% on short–term rates and 7% – 9% for longer maturities.

Furthermore, other two unstable situations occurred in the period November 2014 – July 2015 and, more recently, from January 2022 to June 2022, with rates increments in the range 7% – 12% and 4% – 8% respectively across all maturities. This spiky behavior sinks its roots in both the 2014 and 2022 crisis which opposed the Russian

Federation and the Western countries and brought to economic and trade sanctions, combined to the weakening of the Russian National currency. In both the occasions, the Central Bank of the Russian Federation increased interest rates and adopted policies like inflation targeting and the floating exchange rate regime to support the banking sector and the national currency against external shocks. Such periods of political and economic tension led, as a consequence, to extreme yield curves behaviour. This is evident from Figure 2.2 *(b)*. Here we illustrate yield curves at various times highlighting in red the flat and the inverted behavior that occurred in correspondence of the above described events.

Considering, *ceteris paribus*, the Euro Zone case illustrated in Figure 2.1 *(f)* and 2.3, it is possible to pinpoint that spot rates across the maturity spectrum are less volatile and that the level of spot rates is 3 to 10 times lower: peaks do not exceed the 5% threshold or in some occasion are even negative. These facts indicate an overall higher resistance to exogenous shocks due to higher confidence of the markets and its participants in the macro–economic and financial structure and stability of the EU.

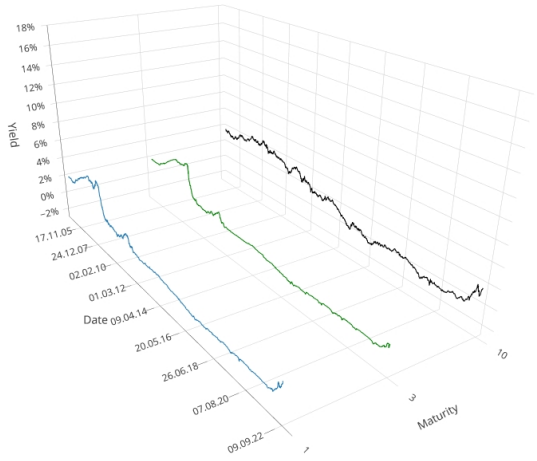


Figure 2.3 *– Euro Zone daily rates setting the maturity to 1, 3, and 10 years.*

In the light of all the above, it clearly emerges that economic and geopolitical crises exert significant pressure on bond markets of emerging countries which seem to be less "immune" to shocks than more developed economies.

As financial turmoils are frequent events in emerging markets, we can reasonally extend to other markets in the sample the remarks discussed in the case of Russia.

2.2.2 Models

In this subsection we describe the three and five–factor parametric models employed in our study.

The Nelson–Siegel model (3F–NS) is a parsimonious three–factor parametric model which has proved to capture a wide range of monotonic, humped and S–type yield curve shapes (De Pooter, 2007).

Let us consider a Zero–Coupon Bond (ZCB) and denote by $y(t,m)$ the observable rate at time $t = 1,2,\ldots,T$, where T is the number of available observations, and maturity $m \in M = (m_1, m_2, \ldots, m_N)'$ representing either a fraction or a multiple of the year, with N being the maximum number of examined maturities. Following Nelson and Siegel (1987), the interpolated spot value at time t can be represented via the parametric function:

$$NS(t,m,\boldsymbol{\beta},\lambda) = \beta_0 + \beta_1 \left(\frac{1-e^{-\lambda m}}{\lambda m} \right) + \beta_2 \left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \right) + \eta_m \qquad (2.1)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is the parameters vector whose components represent, respectively, the impact of the constant long–term component (β_0) that moves the curve up or down; the contribution of the short–term component (β_1) controlling the curve slope, and the effect of the medium–term component (β_2), ruling out the magnitude and direction of the yield curve curvature. The model includes also a decay term λ which controls the convergence speed of the exponential components and determines the position of the peak of the medium–term element, while $\eta_m \sim \mathcal{N}(0, \sigma_m^2)$ is the normally distributed error term with $cov(\eta_r, \eta_s) = 0$, for all $r, s = m_1, \ldots, m_N$, $r \neq s$.

A proper calibration of $\boldsymbol{\beta}$ and λ makes possible an effective replication of a wide variety of yield curve shapes. Parameters estimation is the result of a two–step procedure, where grid search methods (Nelson and Siegel, 1987, Muthoni et al., 2015) identify the value of λ that maximizes the medium–term component, varying the maturity. For each λ the vector of parameters $\hat{\boldsymbol{\beta}}$ is then estimated through an OLS regression choosing in the end the values λ^* and $\hat{\boldsymbol{\beta}}^*$ associated to the highest coefficient of determination.

Later, De Rezende and Ferreira (2008) introduced a five–factor variant (5F–RF) aimed at increasing the flexibility of previous models with additional parameters and decay factors to capture a wider variety of trends. The 5F–RF model, in fact, extends (2.1) including additional short and medium–term components characterized by different decaying factors, to ensure a faster decaying rate and to increase the fitting

ability of yield curve shapes in presence of multiple short–term maxima/minima:

$$RF(t,m,\boldsymbol{\beta},\boldsymbol{\tau}) = \beta_0 + \beta_1\left(\frac{1-e^{-m/\tau_1}}{m/\tau_1}\right) + \beta_2\left(\frac{1-e^{-m/\tau_2}}{m/\tau_2}\right) + \\ + \beta_3\left(\frac{1-e^{-m/\tau_1}}{m/\tau_1} - e^{-m/\tau_1}\right) + \beta_4\left(\frac{1-e^{-m/\tau_2}}{m/\tau_2} - e^{-m/\tau_2}\right) + \eta_m \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$ and $\boldsymbol{\tau} = (\tau_1, \tau_2)'$ are the 5–dimension parameters vector and the decay terms vector, respectively. The estimation of both $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ is based on a two–step procedure that at first identifies the optimal decay parameters $\hat{\tau}_1$, and $\hat{\tau}_2$ in the space Ω of admissible values, by minimizing the Root Mean Square Error (RMSE):

$$\hat{\boldsymbol{\tau}} = (\hat{\tau}_1\ , \hat{\tau}_2)' = \underset{(\tau_1,\ \tau_2)\ \in\ \Omega}{arg\ min} \left\{ \frac{1}{N}\sum_{n=1}^{N} \sqrt{\frac{1}{T}\sum_{t=1}^{T}[y(t,m_n) - \widehat{RF}(t,m_n,\boldsymbol{\tau},\hat{\boldsymbol{\beta}}_t)]^2} \right\} \tag{2.3}$$

where N is the overall number of examined maturities, and T is the number of available observations. Once obtained the optimal vector $\hat{\boldsymbol{\tau}}$, the estimation of the parameters vector $\hat{\boldsymbol{\beta}}$ takes place by applying the OLS regression for each time t.

2.2.3 Optimal Parameters Estimation

As seen in the previous section, the parameters λ and $\boldsymbol{\tau} = (\tau_1, \tau_2)'$ play a fundamental role in the 3F–NS and 5F–RF models respectively, because they drive the decay rate of the exponential components controlling the trend dynamics of the fitted yield curves. The choice of λ and $\boldsymbol{\tau}$ generates a trade–off in the fitting accuracy at both the left and right–handed tails of the yield curve. In fact, small values of λ (big values of τ_1, τ_2) lead to a slow decay of the curve, and hence assure a better fit at longer maturities, but the same is not true at short maturities, especially in presence of sudden and marked curvatures. Conversely, higher values of λ (small values of τ_1, τ_2) result in a quick decay and hence a better fit at short maturities, with an accuracy loss in the long run.

Managing the decay parameters is therefore of paramount importance, as testified by the solutions suggested in the literature. A common approach consists in setting them to the value that maximises the curvature factor at the maturity m where humps or basins are empirically observed. For example, working on U.S. Treasury data, Diebold and Li (2006) and De Pooter (2007) set $\hat{\lambda} = 0.0609$ ($m = 30$ months), while Diebold et al. (2006) assumed $\hat{\lambda} = 0.077$ ($m = 23.3$ months). Moreover, Muvingi and Kwinjo (2014) assumed $\hat{\lambda} = 0.25$ ($m = 7$ months) for the Bank of Zimbabwe certificates, while De Rezende and Ferreira (2008), based on ID-PRE Swap data of the Brazilian

market, set $\hat{\tau}_1$ and $\hat{\tau}_2$ at the best of the estimated values according to (2.3). This latter approach simplifies the numerical optimization process as it linearizes the estimation process of the model with the use of the least-squares regression; in addition, it seems to reach a good compromise between the long and short–run accuracy issues. However assuming constant decay terms is in conflict with the evidence that the term structure of interest rates may show time changes of different intensity in terms of both slope and curvature as observed for all the BRICS (see Figure 2.1): an *a priori* selection of either λ or $\boldsymbol{\tau}$ can therefore lead to non–optimal estimations, weakening the models fitting ability.

For the above reasons, we adopted a different approach, and we considered the decay components as time–varying parameters as well. We applied a two–step estimation procedure to determine the proper values $\lambda^*(t)$, $\tau_1^*(t)$, $\tau_2^*(t)$ and $\hat{\boldsymbol{\beta}}^*(t)$ such that the complete sets of parameters in both the three and five factor cases (herein after named 3F–DNS and 5F–DRF respectively) are at the best for each time t. For the 3F-DNS model, the algorithm works as follows

Step 1: For each market define the set $\boldsymbol{M} = \{m_k\}_{k=1,\dots,N}$ of maturities m_k with N equal to the sets cardinality. In particular, $m_1 = m_L$ is the lower bound of $\boldsymbol{M}$ and corresponds to the first available maturity in the market, while the upper bound m_U is the longest observed maturity. Values in $\boldsymbol{M}$ ranges between corresponding lower/upper values by proper step size Δ.

Step 2: For each $m_k \in \boldsymbol{M}$ with $k = 1, \dots, N$, estimate at time t the value $\lambda_k^*(t)$ that maximizes the curvature term component:

$$\frac{1 - e^{-\lambda(t)m_k}}{\lambda(t)m_k} - e^{-\lambda(t)m_k}, \qquad k = 1, \dots, N$$

In this way, at each time t it is possible to associate the array $\hat{\boldsymbol{\lambda}}(t) = \{\lambda_k^*(t)\}_{k=1,\dots,N}$.

Step 3: For each time $t = 1, \dots, T$:

i) use the array $\hat{\boldsymbol{\lambda}}(t)$ found in Step 2 to estimate the parameters vector $\hat{\boldsymbol{\beta}}(t)$ via the OLS regression;

ii) choose the vector $\hat{\boldsymbol{\beta}}^*(t)$ associated to the lowest value of the Sum of Squared Residuals (SSR):

$$SSR(t) = \sum_{k=1}^{N} [y(t, m_k) - \widehat{DNS}(t, m_k, \hat{\boldsymbol{\beta}}(t), \lambda_k^*(t)]^2$$

iii) repeat steps (i) – (ii) for each time t $(t = 1, \ldots, T)$ to get the time series of the parameter $\lambda^*(t)$.

For what it concerns the 5F–DRF model the estimation procedure is similar to that discussed for the 3F-DNS, although a bit more tricky, due to the presence of two decay factors (τ_1 and τ_2) at each time instead that only one. In this case, the procedure works as follows:

Step 1: For each market define the set $\boldsymbol{M}_j = \{m_{j,k}\}_{k=1,\ldots,N_j}$ of maturities $m_{j,k}$ with $j = 1, 2$ and N_j equal to the sets cardinality; $m_{1,1} = m_{1,L}$ represents the lower bound of $\boldsymbol{M}_1$ and corresponds to the first available maturity of the market, while the upper bound $m_{1,U}$ is, at the same time, the lower bound of $\boldsymbol{M}_2$ $(m_{1,U} = m_{2,L})$ and it is equal to the straddling maturity between the short and medium–term period. Finally, the upper bound of $\boldsymbol{M}_2$ $(m_{2,U})$ is the longest observed maturity. As above, values in $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ range between corresponding lower/upper values by proper step sizes Δ_1 and Δ_2.

Step 2: For each $m_{j,k} \in \boldsymbol{M}_j$ with $k = 1, \ldots, N_j$ and $j = 1, 2$ estimate the vectors $\hat{\boldsymbol{\tau}}_{\mathbf{1}}(t)$ and $\hat{\boldsymbol{\tau}}_{\mathbf{2}}(t)$ that maximize the curvature term components:

$$\frac{1 - e^{-m_{j,k}/\tau_j(t)}}{m_{j,k}/\tau_j(t)} - e^{-m_{j,k}/\tau_j(t)}, \qquad k = 1, \ldots, N$$

Step 3: For every $t = 1, \ldots, T$:

i) for each component of $\hat{\boldsymbol{\tau}}_{\mathbf{1}}$, vary the components of $\hat{\boldsymbol{\tau}}_{\mathbf{2}}$ to estimate by OLS regression different array sets $\hat{\boldsymbol{\beta}}(t)$ choosing the one with the lowest Sum of Squared Residuals (SSR) computed as the squared difference between the observed and estimated rates:

$$SSR(t) = \sum_{n=1}^{N} [y(t, m_k) - \widehat{DRF}(t, m_k, \hat{\boldsymbol{\beta}}(t), \boldsymbol{\tau}(t))]^2$$

Clearly there are as many sets of optimal parameters as the number of $\hat{\boldsymbol{\tau}}_{\mathbf{1}}$ components;

ii) choose the optimal set of parameters $\hat{\boldsymbol{\beta}}^*(t)$ associated to the lowest SSR.

iii) repeat steps (i) – (ii) for each time t $(t = 1, \ldots, T)$ to get the time series parameters of both $\tau_1^*(t)$ and $\tau_2^*(t)$.

2.3 Discussion of the fitting results

The study was carried on using the routines of the R package **DeRezende.Ferreira** (Castello and Resta, 2019), developed by the authors and freely available at the Comprehensive R Archive Network (CRAN) repository (https://cran.r-project.org/web/packages/DeRezende.Ferreira/index.html).

2.3.1 Estimation results for the decay factors

To highlight the advantages of optimal time-varying decay parameters, Table 2.1 for the 3F–DNS and Table 2.2 for the 5F–DRF compare for each country the Average Coefficient of Determination (R^2) obtained according to three distinct approaches: (i) employing the optimal values computed through our estimation procedure; (ii) using constant decay terms as suggested in Diebold and Li (2006) and in De Rezende and Ferreira (2008) for the three and five factor models, respectively; (iii) using constant parameters equal to the optimal decay parameters average values.

Table 2.1 *– Comparison of the average R^2 in the 3F–DNS model associated to the estimation of λ. When referring to the average λ value, this latter is given within round brackets.*

	R^2		
Country	**Optimal $\lambda^*(t)$ Values**	**Constant Value $\lambda = 0.0609$**	**Average Value of $\lambda^*(t)$**
Brazil	0.9899	0.9784	0.9779 ($\lambda = 0.6847$)
Russia	0.9909	0.9474	0.9556 ($\lambda = 0.5049$)
India	0.9738	0.9540	0.9419 ($\lambda = 0.4616$)
China	0.9819	0.9094	0.9707 ($\lambda = 0.2190$)
South Africa	0.9959	0.9607	0.9518 ($\lambda = 0.2539$)

The choice of optimal decay terms (second column in Tables 2.1 and 2.2) increase the degree of fitting accuracy of both models for every BRICS country: the R^2 values of the 3F-DNS model are on average 3.90% better than those obtained with constant λ (the lowest increase is 1.18% for Brazil and the highest is 7.97% for China). Furthermore the results with our technique are also higher than those obtained with λ average values: the appreciation of the R^2 ranges from 1.15% for China to 4.63% for South Africa. Similar remarks hold also when we turn on Table 2.2 which considers the

5F-DRF model. In this case the average improvement with respect to keep τ_1 and τ_2 constant is 1.58%, with the minimum (+0.05%) for South Africa and the maximum (+4.61%) for Russia. In addition, when comparing the results in columns 1 and 3 we observe that time–varying parameters $\tau_1(t)$ and $\tau_2(t)$ result in higher R^2 values with an average increase of +1.56%; the minimum increase (+0.06%) is associated to South Africa and the highest (+4.61%) is recorded for Russia.

Table 2.2 *– Comparison of the average R^2 for the 5F–DRF model using optimal $\boldsymbol{\tau_1^*(t)}$, $\boldsymbol{\tau_2^*(t)}$ values (second column), constant values (third column) and average values (fourth column). Average values of τ_1 and τ_2 are given within brackets*

	R^2		
Country	**Optimal $\tau_1^*(t)$ and $\tau_2^*(t)$ Values**	**Constant Values**	**Average Values of $\tau_1^*(t)$ and $\tau_2^*(t)$**
Brazil	0.9999	0.9779 ($\tau_1 = 1.115$ $\tau_2 = 4.182$)	0.9779 ($\tau_1 = 0.775$ $\tau_2 = 3.561$)
Russia	0.9997	0.9556 ($\tau_1 = 0.976$ $\tau_2 = 13.941$)	0.9556 ($\tau_1 = 1.073$ $\tau_2 = 6.022$)
India	0.9971	0.9896 ($\tau_1 = 1.394$ $\tau_2 = 5.855$)	0.9907 ($\tau_1 = 1.841$ $\tau_2 = 10.338$)
China	0.9982	0.9957 ($\tau_1 = 3.067$ $\tau_2 = 14.777$)	0.9958 ($\tau_1 = 2.620$ $\tau_2 = 11.799$)
South Africa	0.9999	0.9994 ($\tau_1 = 1.812$ $\tau_2 = 14.220$)	0.9993 ($\tau_1 = 2.416$ $\tau_2 = 9.220$)

To gain a better intuition of these results in Figures 2.4 – 2.5 we compared the average MSE generated by each model according to the approaches (i) to (iii).

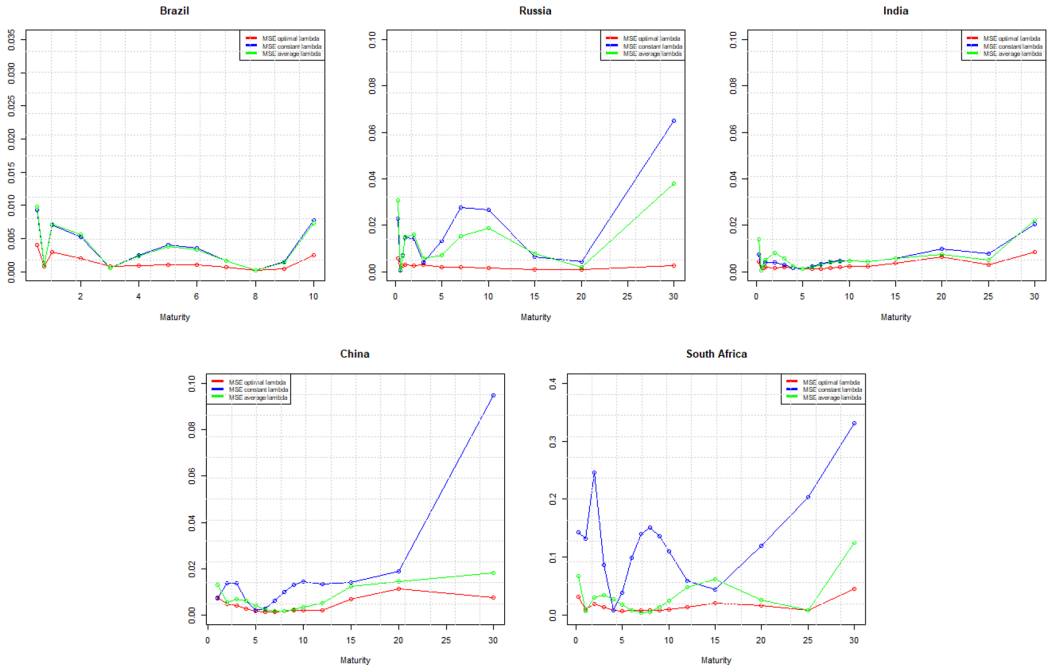


Figure 2.4 – *MSE generated using time–variant λ (red), constant λ (blue) and average λ (green).*

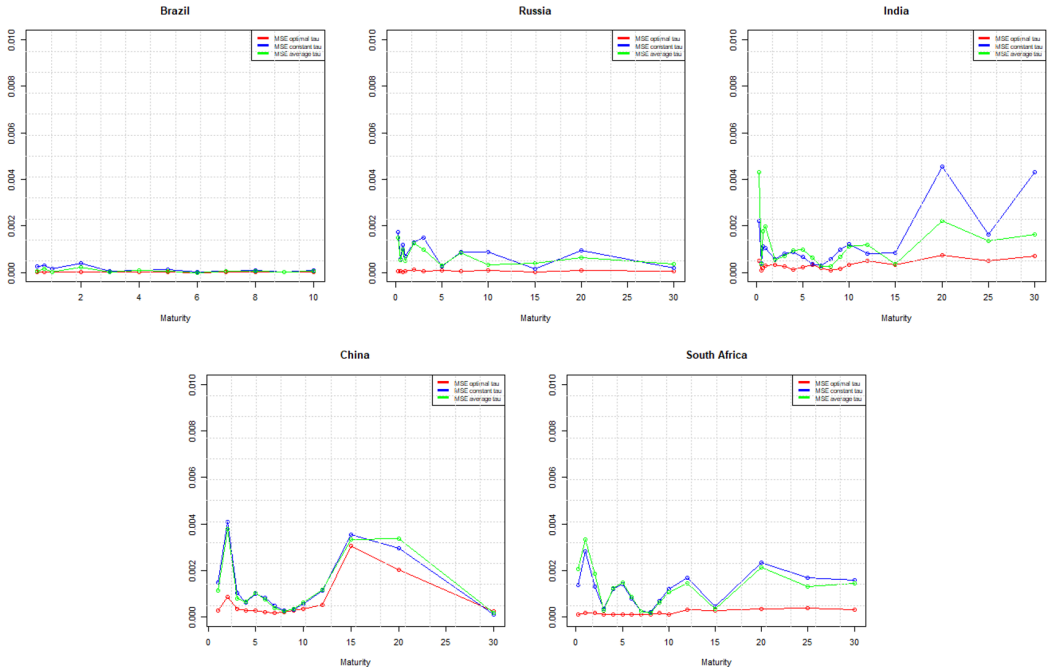


Figure 2.5 – *MSE generated using time–variant τ (red), constant τ (blue) and average τ (green).*

We can preliminary state that time–variant λ and $\boldsymbol{\tau}$ bring additional flexibility to the examined models, with the improvement of the fitting as confirmed by lower error values. On the contrary, constant λ (τ_1, τ_2) values make harder the fitting especially with much cumbersome curves like those of the BRICS markets.

2.3.2 In–Sample fitting performance analysis

The results of the comparison between the three and five factor models are firstly presented in terms of average fitted spot rates per maturity, plotted in Figure 2.6 for each country and model.

Overall, the 3F–DNS (blue) and 5F–DRF (red) estimated curves are almost perfectly matching to the observed (black) ones. The 3F–DNS model, however, exhibits some overestimation issues in correspondence of the medium–term maturities as can be seen in the Russian and Chinese cases with the estimated curve lying slightly above the observed one, and underestimation in the long term which are particularly evident in the Indian, Chinese and South African markets with the blue curve going under the observed one; on the contrary when using the 5F–DRF model a slight over (under)–estimation is present in the middle (final) section of the curve only in the case of the Indian market.

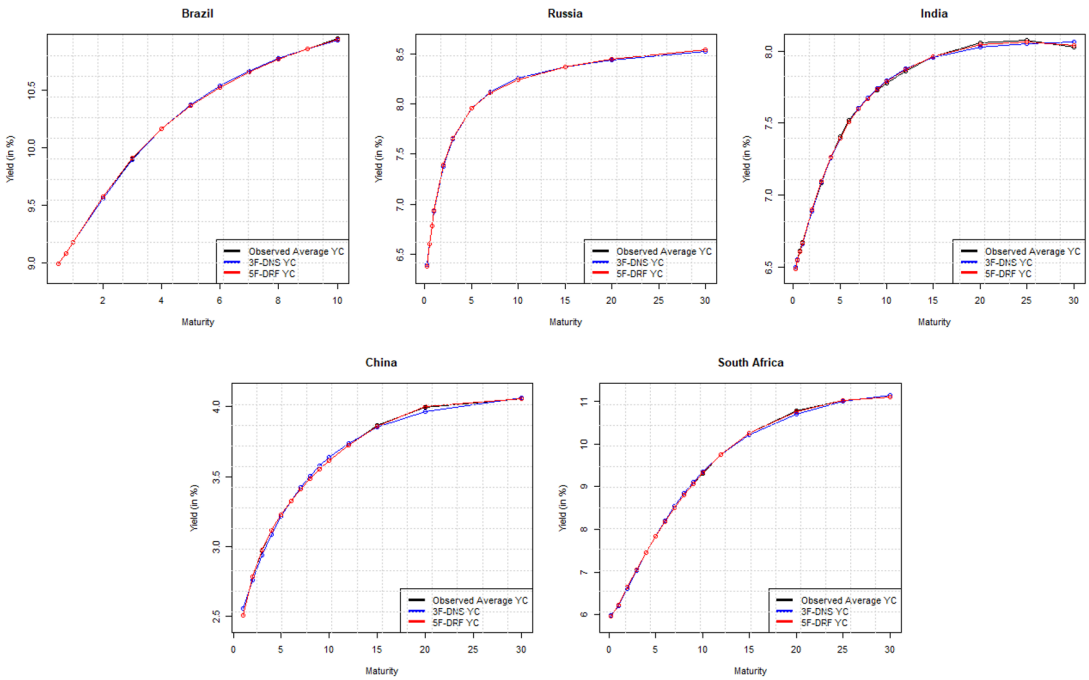


Figure 2.6 *– Comparison of average observable yield curves (black) with average yield estimates obtained with the 3F-DNS (blue) and 5F-DRF (red) models.*

We have also calculated the Average Squared Error, given as the difference between the observed and fitted average yield curve values. Both models performed very well: error values span from a minimum of 7.01×10^{-5} (Brazil) to a maximum of 1.02×10^{-3} (South Africa) for the 3F-DNS model, and from a minimum of 2.15×10^{-7} (Russia) to a maximum of 6.80×10^{-5} (India) for the 5F-DRF model.

Furthermore we compared the values of some well–known goodness of fit indicators: R^2, MSE and $RMSE$ whose main statistics are shown in Table 2.3. The R^2 was obtained as the result of the parameters estimation process via the Ordinary Least–Squares (OLS) method. Using the residuals generated by the OLS we computed the MSE and $RMSE$ metrics for each yield curve as well, hence obtaining the related time series which are plotted in Appendix 2.8. In order to give a more granular view of the models performances, we determined the average MSE and $RMSE$ generated by the 3F–DNS and 5F–DRF models at each maturity for every country summarizing the results in Table 2.4.

Both models generated quite low error values; nevertheless, deepening the analysis and looking at the error metrics time series of Figure 2.8 the 3F–DNS model presents higher error peaks in all the examined countries, mainly during turbulent periods. This is probably due to the well–known difficulties of the 3F–DNS model to fit more dynamic behavior (De Pooter, 2007), i.e. twisted and/or humped shapes.

Based on the data of Table 2.3 and Table 2.4 we show that the five factor model generated the lowest MSE and $RMSE$ values, achieving a significant improvement over the 3F–DNS model as well: in fact for each country we can observe a reduction of the MSE ranging between a minimum of 84.34% and a maximum of 98.99% for China and Brazil, respectively; if we consider the average results for each country and maturity we detect a reduction of the MSE by a factor of 10 to 1000.

Overall, the 3F–DNS and 5F–DRF models ensured satisfying fitting and they were able to preserve the shapes avoiding unreasonable under/over estimations. The five–factor model, however, produced more accurate approximations of the observed curves in almost four countries out of five.

In the light of the above results it is possible to state that both models with time–varying parameters are adequate tools for yield curve modeling even when applied to developing countries whose bond market has more complex dynamics. However, the 5F-DRF performs better since it benefits of improved flexibility due to both time–variant decay terms and additional slope and curvature parameters.

Table 2.3 *– Comparison of R^2, MSE and RMSE for the 3F-DNS and 5F-DRF models. For each model and country Mean, Standard Deviation (SD), Minimum (Min) and Maximum (Max) value of the metrics are reported.*

		R^2		MSE		RMSE	
		3F-DNS	**5F-DRF**	**3F-DNS**	**5F-DRF**	**3F-DNS**	**5F-DRF**
Brazil	***Mean***	0.98988	0.99991	1.4355×10^{-3}	1.4486×10^{-5}	2.8805×10^{-2}	2.9619×10^{-3}
	SD	0.04979	0.00062	2.6952×10^{-3}	6.2670×10^{-5}	2.4619×10^{-2}	2.3908×10^{-3}
	Min	0.14671	0.97458	1.1615×10^{-6}	1.4523×10^{-8}	1.0778×10^{-3}	1.2051×10^{-4}
	Max	0.99999	1	4.2524×10^{-2}	2.8513×10^{-3}	2.0621×10^{-1}	5.3398×10^{-2}
Russia	***Mean***	0.99088	0.99965	2.3245×10^{-3}	6.6275×10^{-5}	3.3847×10^{-2}	6.0255×10^{-3}
	SD	0.03863	0.00249	6.9326×10^{-3}	1.5886×10^{-4}	3.4338×10^{-2}	5.4749×10^{-3}
	Min	0.26261	0.93522	1.9335×10^{-6}	4.5277×10^{-7}	1.3905×10^{-3}	6.7288×10^{-4}
	Max	0.99999	1	1.3340×10^{-1}	2.7742×10^{-3}	3.6524×10^{-1}	5.2671×10^{-2}
India	***Mean***	0.97376	0.99714	2.6349×10^{-3}	3.3138×10^{-4}	4.4275×10^{-2}	1.4651×10^{-2}
	SD	0.04678	0.00876	3.4680×10^{-3}	5.5550×10^{-4}	2.5978×10^{-2}	1.0806×10^{-2}
	Min	0.29763	0.71665	9.4046×10^{-6}	5.3692×10^{-7}	3.0667×10^{-3}	7.3275×10^{-4}
	Max	0.99994	1	4.4733×10^{-2}	8.6687×10^{-3}	2.1150×10^{-1}	9.3106×10^{-2}
China	***Mean***	0.98186	0.99815	4.1556×10^{-3}	6.5080×10^{-4}	5.1737×10^{-2}	1.3578×10^{-2}
	SD	0.03201	0.00761	7.8732×10^{-3}	2.8618×10^{-3}	3.8461×10^{-2}	2.1599×10^{-2}
	Min	0.22041	0.79261	6.5712×10^{-6}	4.5227×10^{-8}	2.5634×10^{-3}	2.1267×10^{-4}
	Max	0.99997	1	1.4055×10^{-1}	6.3573×10^{-2}	3.7489×10^{-1}	2.5214×10^{-1}
South Africa	***Mean***	0.99594	0.99988	1.5003×10^{-2}	1.9223×10^{-4}	8.8286×10^{-2}	9.8688×10^{-3}
	SD	0.00494	0.00123	5.3505×10^{-2}	1.5302×10^{-3}	8.4921×10^{-2}	9.7402×10^{-3}
	Min	0.92787	0.94945	2.7715×10^{-6}	2.4612×10^{-7}	1.6648×10^{-3}	4.9611×10^{-4}
	Max	0.99999	1	1.1823	6.1278×10^{-2}	1.0873	2.4754×10^{-1}

Table 2.4 *– Comparison of the average MSE and RMSE generated by the 3F-DNS and 5F-DRF models for each countries maturity.*

Country	*Brazil*				*Russia*				*India*				*China*				*South Africa*			
Model	*3F-DNS*		*5F-DRF*		*3F-DNS*		*5F-DRF*		*3F-DNS*		*5F-DRF*		*3F-DNS*		*5F-DRF*		*3F-DNS*		*5F-DRF*	
Maturity	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE
0.25	-	-	-	-	5.9×10^{-3}	7.7×10^{-2}	4.3×10^{-5}	6.6×10^{-3}	4.4×10^{-3}	6.7×10^{-2}	4.9×10^{-4}	2.2×10^{-2}	-	-	-	-	3.2×10^{-2}	1.8×10^{-1}	1.0×10^{-4}	1.0×10^{-2}
0.5	4.0×10^{-3}	6.3×10^{-2}	8.9×10^{-6}	2.9×10^{-3}	8.8×10^{-4}	2.9×10^{-2}	5.7×10^{-5}	7.6×10^{-3}	7.0×10^{-4}	2.6×10^{-2}	1.0×10^{-4}	1.0×10^{-2}	-	-	-	-	-	-	-	-
0.75	7.7×10^{-4}	2.8×10^{-2}	3.1×10^{-5}	5.6×10^{-3}	2.6×10^{-3}	5.1×10^{-2}	3.8×10^{-5}	6.1×10^{-3}	1.5×10^{-3}	3.9×10^{-2}	2.0×10^{-4}	1.4×10^{-2}	-	-	-	-	-	-	-	-
1	2.9×10^{-3}	5.4×10^{-2}	1.8×10^{-5}	4.2×10^{-3}	3.0×10^{-3}	5.5×10^{-2}	5.1×10^{-5}	7.1×10^{-3}	1.8×10^{-3}	4.3×10^{-2}	3.0×10^{-4}	1.7×10^{-2}	7.7×10^{-3}	8.8×10^{-2}	2.7×10^{-4}	1.6×10^{-2}	1.0×10^{-2}	1.0×10^{-1}	1.9×10^{-4}	1.4×10^{-2}
2	1.9×10^{-3}	4.4×10^{-2}	2.5×10^{-5}	5.0×10^{-3}	2.6×10^{-3}	5.1×10^{-2}	1.3×10^{-4}	1.1×10^{-2}	1.8×10^{-3}	4.2×10^{-2}	3.2×10^{-4}	1.8×10^{-2}	4.9×10^{-3}	6.9×10^{-2}	8.7×10^{-4}	2.9×10^{-2}	2.0×10^{-2}	1.4×10^{-1}	1.9×10^{-4}	1.4×10^{-2}
3	7.6×10^{-4}	2.8×10^{-2}	2.0×10^{-5}	4.5×10^{-3}	2.9×10^{-3}	5.3×10^{-2}	7.0×10^{-5}	8.3×10^{-3}	2.1×10^{-3}	4.5×10^{-2}	2.5×10^{-4}	1.6×10^{-2}	4.3×10^{-3}	6.6×10^{-2}	3.4×10^{-4}	1.8×10^{-2}	1.4×10^{-2}	1.2×10^{-1}	1.2×10^{-4}	1.1×10^{-2}
4	8.9×10^{-4}	2.9×10^{-2}	1.3×10^{-5}	3.6×10^{-3}	-	-	-	-	1.5×10^{-3}	3.9×10^{-2}	1.4×10^{-4}	1.2×10^{-2}	2.9×10^{-3}	5.7×10^{-2}	3.0×10^{-4}	1.7×10^{-2}	9.1×10^{-3}	9.5×10^{-2}	1.1×10^{-4}	1.0×10^{-2}
5	1.1×10^{-3}	3.3×10^{-2}	9.8×10^{-6}	3.1×10^{-3}	2.1×10^{-3}	4.6×10^{-2}	7.8×10^{-5}	8.8×10^{-3}	1.2×10^{-3}	3.4×10^{-2}	2.4×10^{-4}	1.6×10^{-2}	1.9×10^{-3}	4.3×10^{-2}	2.7×10^{-4}	1.7×10^{-2}	7.8×10^{-3}	8.8×10^{-2}	1.0×10^{-4}	1.0×10^{-2}
6	1.1×10^{-3}	3.3×10^{-2}	6.6×10^{-6}	2.6×10^{-3}	-	-	-	-	1.2×10^{-3}	3.4×10^{-2}	3.2×10^{-4}	1.8×10^{-2}	1.3×10^{-3}	3.6×10^{-2}	2.1×10^{-4}	1.5×10^{-2}	8.0×10^{-3}	8.9×10^{-2}	1.1×10^{-4}	1.0×10^{-2}
7	6.9×10^{-4}	2.6×10^{-2}	1.3×10^{-5}	3.6×10^{-3}	1.8×10^{-3}	4.3×10^{-2}	6.8×10^{-5}	8.2×10^{-3}	1.4×10^{-3}	3.8×10^{-2}	2.1×10^{-4}	1.4×10^{-2}	1.3×10^{-3}	3.7×10^{-2}	1.9×10^{-4}	1.4×10^{-2}	8.1×10^{-3}	9.0×10^{-2}	1.0×10^{-4}	1.0×10^{-2}
8	1.9×10^{-4}	1.4×10^{-2}	8.9×10^{-6}	2.9×10^{-3}	-	-	-	-	1.8×10^{-3}	4.2×10^{-2}	1.1×10^{-4}	1.0×10^{-2}	1.7×10^{-3}	4.1×10^{-2}	2.1×10^{-4}	1.5×10^{-2}	8.3×10^{-3}	9.1×10^{-2}	1.0×10^{-4}	1.0×10^{-2}
9	4.1×10^{-4}	2.0×10^{-2}	9.1×10^{-6}	3.0×10^{-3}	-	-	-	-	2.1×10^{-3}	4.5×10^{-2}	1.6×10^{-4}	1.3×10^{-2}	1.9×10^{-3}	4.4×10^{-2}	2.7×10^{-4}	1.6×10^{-2}	8.7×10^{-3}	9.3×10^{-2}	1.8×10^{-4}	1.4×10^{-2}
10	2.4×10^{-3}	4.9×10^{-2}	1.1×10^{-5}	3.3×10^{-3}	1.8×10^{-3}	4.2×10^{-2}	1.0×10^{-4}	1.0×10^{-2}	2.2×10^{-3}	4.7×10^{-2}	3.2×10^{-4}	1.8×10^{-2}	2.2×10^{-3}	4.6×10^{-2}	3.4×10^{-4}	1.8×10^{-2}	9.7×10^{-2}	9.8×10^{-1}	1.1×10^{-4}	1.0×10^{-2}
12	-	-	-	-	-	-	-	-	2.3×10^{-3}	4.8×10^{-2}	5.0×10^{-4}	2.2×10^{-2}	2.3×10^{-3}	4.8×10^{-2}	5.2×10^{-4}	2.3×10^{-2}	1.4×10^{-2}	1.2×10^{-1}	3.2×10^{-4}	1.8×10^{-2}
15	-	-	-	-	9.8×10^{-4}	3.1×10^{-2}	3.9×10^{-5}	6.2×10^{-3}	3.6×10^{-3}	5.9×10^{-2}	3.3×10^{-4}	1.8×10^{-2}	6.9×10^{-3}	8.3×10^{-2}	3.1×10^{-3}	5.5×10^{-2}	2.0×10^{-2}	1.4×10^{-1}	3.0×10^{-4}	1.7×10^{-2}
20	-	-	-	-	7.9×10^{-4}	2.8×10^{-2}	7.7×10^{-5}	8.8×10^{-3}	6.5×10^{-3}	8.1×10^{-2}	7.6×10^{-4}	2.8×10^{-2}	1.1×10^{-2}	1.1×10^{-1}	2.0×10^{-3}	4.5×10^{-2}	1.7×10^{-2}	1.3×10^{-1}	3.6×10^{-4}	1.9×10^{-2}
25	-	-	-	-	-	-	-	-	2.9×10^{-3}	5.4×10^{-2}	5.1×10^{-4}	2.3×10^{-2}	-	-	-	-	8.0×10^{-3}	8.9×10^{-2}	3.7×10^{-4}	1.9×10^{-2}
30	-	-	-	-	2.6×10^{-3}	5.1×10^{-2}	4.4×10^{-5}	6.6×10^{-3}	8.5×10^{-3}	9.2×10^{-2}	6.9×10^{-4}	2.6×10^{-2}	7.8×10^{-3}	8.8×10^{-2}	2.4×10^{-4}	1.5×10^{-2}	4.5×10^{-2}	2.1×10^{-1}	3.1×10^{-4}	1.7×10^{-2}

2.4 Forecasting the BRICS Term Structure

2.4.1 Competing models

In addition to in–sample fitting, a term structure model should also provide effective out–of–sample prediction capabilities of the yield curve, we therefore investigated the models predictive abilities.

We carried out the task in a kind of indirect way: instead of directly predicting spot rates we forecasted the building blocks of the yield curve, that is parameters and decay terms. We used various techniques to model and predict the time series of $\hat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\boldsymbol{\tau}}$: the Univariate Autoregressive AR(1) model, the Trigonometric seasonal Box–Cox Transformation with ARMA residuals Trend and Seasonal components (TBATS), and a combination of Autoregressive Integrated Moving Average and Nonlinear Autoregressive Neural Network (ARIMA–NARNN). We then estimated spot rates at the time $t+h$ and maturity $m \in M = (m_1, m_2, \ldots, m_N)'$:

$$\hat{y}_{DNS}(t+h|t) = \hat{\beta}_{0,t+h|t} + \hat{\beta}_{1,t+h|t}\left(\frac{1-e^{-\hat{\lambda}_{t+h|t}\,m}}{\hat{\lambda}_{t+h|t}\,m}\right) + \hat{\beta}_{2,t+h|t}\left(\frac{1-e^{-\hat{\lambda}_{t+h|t}\,m}}{\hat{\lambda}_{t+h|t}\,m} - e^{-\hat{\lambda}_{t+h|t}\,m}\right), \tag{2.4}$$

for the three factor model, and

$$\begin{aligned}\hat{y}_{DRF}(t+h|t) = \hat{\beta}_{0,t+h|t} &+ \hat{\beta}_{1,t+h|t}\left(\frac{1-e^{-m/\hat{\tau}_{1,t+h|t}}}{m/\hat{\tau}_{1,t+h|t}}\right) + \hat{\beta}_{2,t+h|t}\left(\frac{1-e^{-m/\hat{\tau}_{2,t+h|t}}}{m/\hat{\tau}_{2,t+h|t}}\right) + \\ &+ \hat{\beta}_{3,t+h|t}\left(\frac{1-e^{-m/\hat{\tau}_{1,t+h|t}}}{m/\hat{\tau}_{1,t+h|t}} - e^{-m/\hat{\tau}_{1,t+h|t}}\right) + \\ &+ \hat{\beta}_{4,t+h|t}\left(\frac{1-e^{-m/\hat{\tau}_{2,t+h|t}}}{m/\hat{\tau}_{1,t+h|t}} - e^{-m/\hat{\tau}_{2,t+h|t}}\right),\end{aligned} \tag{2.5}$$

for the five factor model.

When using the AR(1) process to predict the parameters in (2.4) and (2.5) we have:

$$x_{k,t+h} = \alpha_0 + \alpha_1 x_{k,t} + \varepsilon_{k,t} \tag{2.6}$$

where $x_{k,t+h}(x_{k,t})$ is the variable to model, i.e. either β, λ or τ_1, τ_2, while α_0 is the coefficient of the zero degree term; α_1 is the coefficient of the autoregressive term and ε_t is a white noise error term with $E(\varepsilon_{k,t}) = 0$ and $Var(\varepsilon_{k,t}) = \sigma_k^2$.

In a similar fashion, when using the TBATS, we have:

$$x_{k,t+h}^{(\omega)} = l_t + \phi b_t + \sum_{j=1}^{T} s_{t+h-\eta_j}^{(i)} + d_{t+h} \tag{2.7}$$

where $x_{k,t+h}^{(\omega)}$ is the Box–Cox transformation of the observations $x_{k,t+h}$ with the Box–Cox parameter ω; l_t and b_t represent, respectively, the local level and the short–run trend at time t; ϕ is the dampening parameter for the trend; $s_{t+h}^{(i)}$ is the i^{th} seasonal component while η_j is the seasonal period; and d_{t+h} is the prediction error modeled as an ARMA(p,q) process.

Finally, the combination of the ARIMA(p,d,q) process and NAR–NN is aimed to provide more flexibility in forecasting the parameters. In fact, the ARIMA(p,d,q) is used to predict each models coefficient $\boldsymbol{\beta}_k$ which is expressed as a linear function of both its past observations and past residual error terms (or random shocks):

$$x_{k,t+h} = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{k,t+h-j} + \sum_{i=1}^{q} \gamma_i \varepsilon_{k,t+h-i} \tag{2.8}$$

with $x_{k,t+h}$ as above, α_0 being the intercept, α_j and γ_i the autoregressive and moving average coefficients respectively, and $\varepsilon_{k,t}$ a white noise process.

On the other hand, the NAR–NN is used to forecast the future values of the decay terms λ and $\boldsymbol{\tau}$ according to the:

$$x_{k,t+h} = \omega_{k,0} + \sum_{i=1}^{n} \omega_{k,i}\ \Lambda\left(\alpha_{i,0} + \sum_{j=1}^{m} \alpha_{i,j} x_{k,t+h-j}\right) + \varepsilon_{k,t} \tag{2.9}$$

where m is the number of lagged input values $x_{k,t+h-j}$; $\alpha_{i,j}$ is the connection weight between the input unit j and the closest hidden unit i; Λ represents the activation function; $\omega_{k,i}$ is the connection weight between the hidden unit i and the output unit k; while $\alpha_{i,0}$ and $\omega_{k,0}$ are the bias used to optimize the working point of the neurons in the hidden and output units respectively; finally $\varepsilon_{k,t}$ represents the error term.

Relatively to the ARIMA model, we followed the Box–Jenkins (Box et al., 2016) and Hyndman–Khandakar (Hyndman and Khandakar, 2008) method to determine the most appropriate (p,d,q) specification. For what concerns the development of the NAR–NNs architecture (i.e. the time delays, number of nodes, hidden layers, activation functions etc.) we followed a trial and error approach due to the absence of specific rules (Lantz, 2019).

2.4.2 Methodology and Performance evaluation

From a practical viewpoint, parameters and hence yield curves prediction was performed on a daily basis in the period June 2022 – September 2022, for an overall number of 50 predicted days for each BRICS country using the sliding window method. Specifically, we chose a quite limited range of values close to the forecasting period on which the models are firstly calibrated and then used for one–step–ahead predictions. After each forecast the window is shifted and updated with a new value in order to predict the next one. The advantage of this approach consists in giving priority to the information content of the period close to that of forecasts since it is intended to deeply influence the events of the near future, thus incorporating the autocorrelation features of the series into the model.

To investigate the predictive performance of the candidate models, we evaluated the statistical accuracy of the forecasts with the Mean Square Forecasting Error (MSFE) and the Mean Absolute Percentage Error (MAPE) performance metrics:

$$MSFE = \frac{1}{n}\sum_{j=1}^{n}(y_{t+h} - \hat{y}_{t+h})^2 \tag{2.10}$$

$$MAPE = \frac{100}{n}\sum_{j=1}^{n}\left|\frac{y_{t+h} - \hat{y}_{t+h}}{y_{t+h}}\right| \tag{2.11}$$

where y_{t+h} is the observed value in $t+h$ and $\hat{y}_{t+h}$ the related forecast.

Main results are summarized in Table 2.5 for each country and method. Furthermore, in Figure 2.7 we compare the average observed yield curves to the average forecasted ones for each country, model and method.

Table 2.5 *– Average MSFE and MAPE (%) metrics obtained with different competing methods.*

		Brazil		**Russia**		**India**		**China**		**S. Africa**	
		3F-DNS	*5F-DRF*	*3F-DNS*	*5F-DRF*	*3F-DNS*	*5F-DRF*	*3F-DNS*	*5F-DRF*	*3F-DNS*	*5F-DRF*
MSFE	AR(1)	0.0635	0.2106	0.0081	0.4775	0.0059	0.1344	0.0034	0.1699	0.0448	0.2298
	TBATS	0.0673	0.4058	0.0240	0.3628	0.0091	0.2235	0.0043	2.1228	0.1675	0.3714
	ARIMA-NARNN	0.1374	0.1449	0.0161	0.3702	0.0062	0.2873	0.0035	0.2439	2.4051	0.1902
MAPE	AR(1)	1.5564	2.5550	0.8015	5.5957	0.8808	3.0490	1.9111	11.4933	1.6075	3.7369
	TBATS	1.5757	3.2530	1.3838	5.1773	1.0227	4.2622	2.0728	32.4487	2.5632	4.0259
	ARIMA-NARNN	2.0695	1.9699	1.1545	5.0771	0.8872	4.7553	1.9255	9.6457	11.6353	3.2622

Data in Table 2.5 can be interpreted in at least two ways. On the one hand, it is possible to detected the most effective forecasting method within each parametric model and market; on the other hand, for each market it is possible to determine which combination of parametric model/forecasting method delivered the overall best results.

Looking at the results of the MAPE indicator over the whole forecasting window we can clearly see the dominance of the 3F–DNS model, which produced overall the best performance delivering accurate predictions over the entire maturity spectrum and across all the countries, with an average accuracy of over 98%.

For what is concerning the 5F–DRF model, the presence of further slope and curvature terms, so important to ensure higher in-sample fitting performances, didn't assure any advantage to the models predictive power with respect to the 3F–DNS. Nevertheless, the 5F–DRF model can effectively replicate the average trends of the BRICS curves; moreover, the small error measures jointly with 95% level of predictive precision makes the 5F–DRF an ideal alternative to the three factor model in every considered market despite the high variability of the parameters.

Going into details and considering the 3F–DNS model, the AR(1) process ensured the best result within all the countries achieving 98.65% of average forecasting accuracy; then comes TBATS with 98,28% and ARIMA–NARNN with 96.47%. Moving to the 5F–DRF model, AR(1) and TBATS achieved the 94.71% and 90.16% predictive accuracy, respectively, while the best result was obtained with the ARIMA–NARNN combination with 95.10% precision. The latter result was attained due to the ability of the neural network to better handle the nonlinear behavior of the decay terms.

Turning our attention to the forecasting combinations and cross–checking the tabulated results, it is possible to highlight that the best predictions were systematically provided by the 3F–DNS–AR(1) combination in every considered country, with an overall average MAPE improvement of 24,52 % with respect to its direct competitors

(3F–DNS–TBATS and 3F–DNS–ARIMA–NARNN), and of 68,26 % with respect to the methods used within the 5F–DRF model.

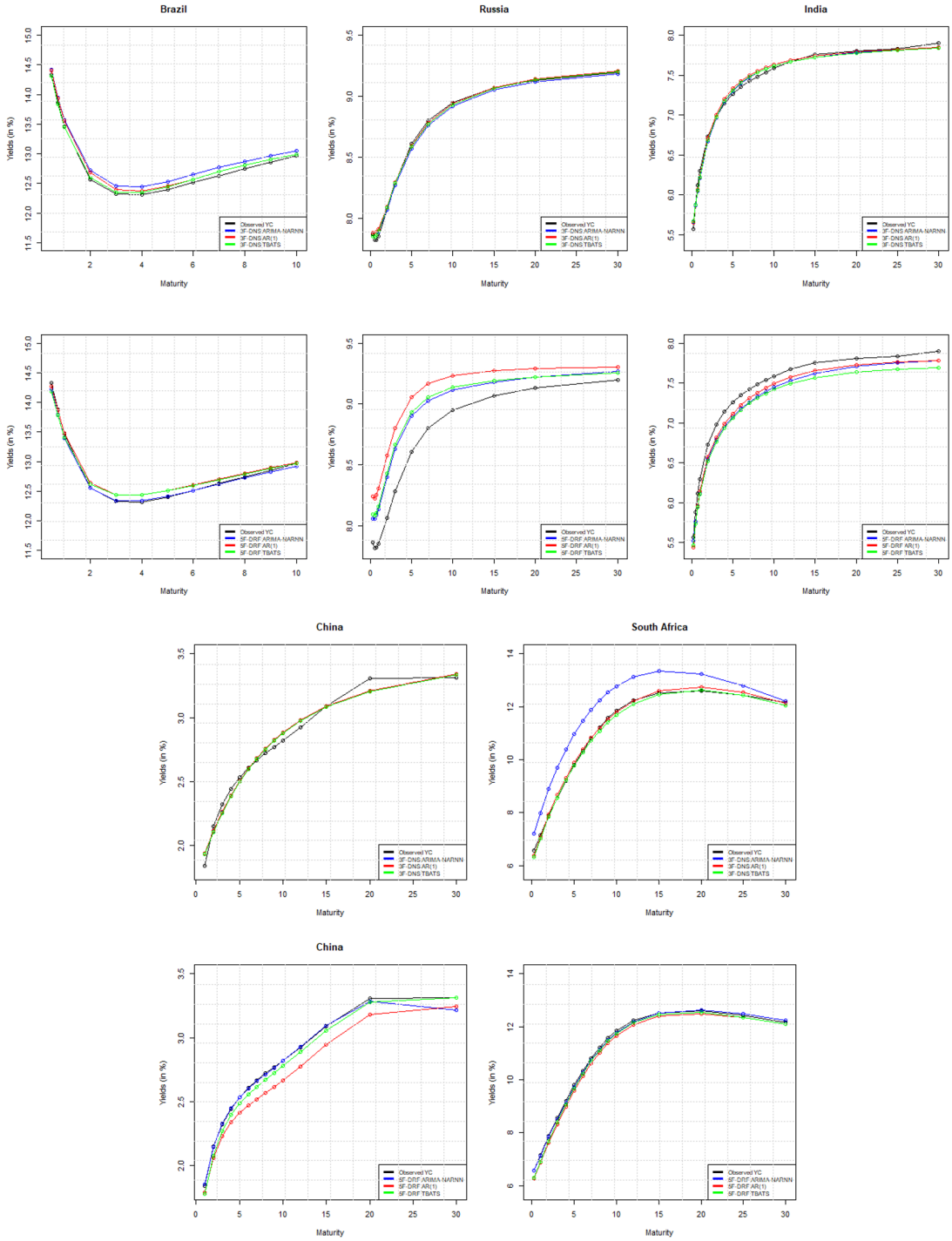


Figure 2.7 *– For each BRICS country (in column) the graph compares yield curves forecasted with different techniques for the 3F-DNS and 5F-DRF models.*

Summarizing the results, it is possible to state that the proposed framework makes it possible to predict with high precision and reliability the challenging dynamics characterising BRICS yield curves avoiding the need to resort to constant decay terms unlike most of similar research. Moreover, the comparison between the two models revealed the existence of a fitting–forecast trade–off: depending on the need, it is possible to opt for the 3F–DNS which ensures more accurate yield curves predictions, but slightly less precise in–sample–fitting; or rely on the 5F-DRF which ensures better fitting abilities but less accurate spot rates predictions.

2.5 Conclusion

In this paper we analyzed a methodology aimed at identifying optimal time–varying parameters for the Three Factor Dynamic Nelson–Siegel (3F–DNS) and the Five Factor Dynamic De Rezende–Ferreira (5F–DRF) models. We tested the modeling and predictive abilities of the proposed framework outside the comfort zone of western economies, that is we focused our attention on BRICS countries. Within the estimation phase we highlighted the advantages of using optimal time–varying decay terms over the constant alternatives. With regard to the predictive process, we moved within the Diebold–Li dynamic framework and employed AR(1), TBATS and a combination of ARIMA–NARNN as Local Data Generating Processes to predict models parameters, and hence yield curves, as an alternative approach to direct interest rates forecasting.

According to the in–sample fitting results, we first found that the use of time–varying decay terms allowed to outperform the results obtained keeping λ and $\boldsymbol{\tau}$ constant or averaging them and ensure the desired flexibility to manage anomalies and extreme dynamics characterizing BRICS markets. Additionally, we showed that both models performed well in–sample as they can describe and replicate the main trends and shapes of BRICS yield curves. However, the 5F–DRF model with multiple decay parameters and additional slope and curvature factors assures improved fitting results. On the contrary the 3F–DNS model generates significantly larger errors due to its well known limitations in approximating the short and long term maturities as well as curves with more inflexion points.

Relatively to the models out–of sample performances, we obtained satisfying results with an average predictive accuracy of over 95%. Overall, the out–of–sample predictions of the 3F–DNS–AR(1) model turned out to be more accurate with lower errors in every market, so that we concluded that not necessarily a richer parametrization ensures also better predictive abilities. The results obtained herein confirm the relevant predictive

power of our approach also within emerging economies without the need to resort to constant decay terms.

The analysis conducted in this paper can be extended in different ways. First of all alternative estimation approaches (e.g. the Maximum Likelihood, Kalman Filter or Machine Learning methods) can be tested, thus avoiding the a priori selection of the decay terms. Eventually, a refined version of the models which integrates financial and macroeconomic factors (e.g. monetary policy, inflation, economic growth) can be considered for the BRICS bond market. Finally it would be interesting to test the proposed framework in different markets (e.g. commodity, derivatives, forex) or for different financial instruments (e.g. corporate bonds, credit default swaps). Actually all these topics represent a part of our ongoing research.

2.6 Appendix

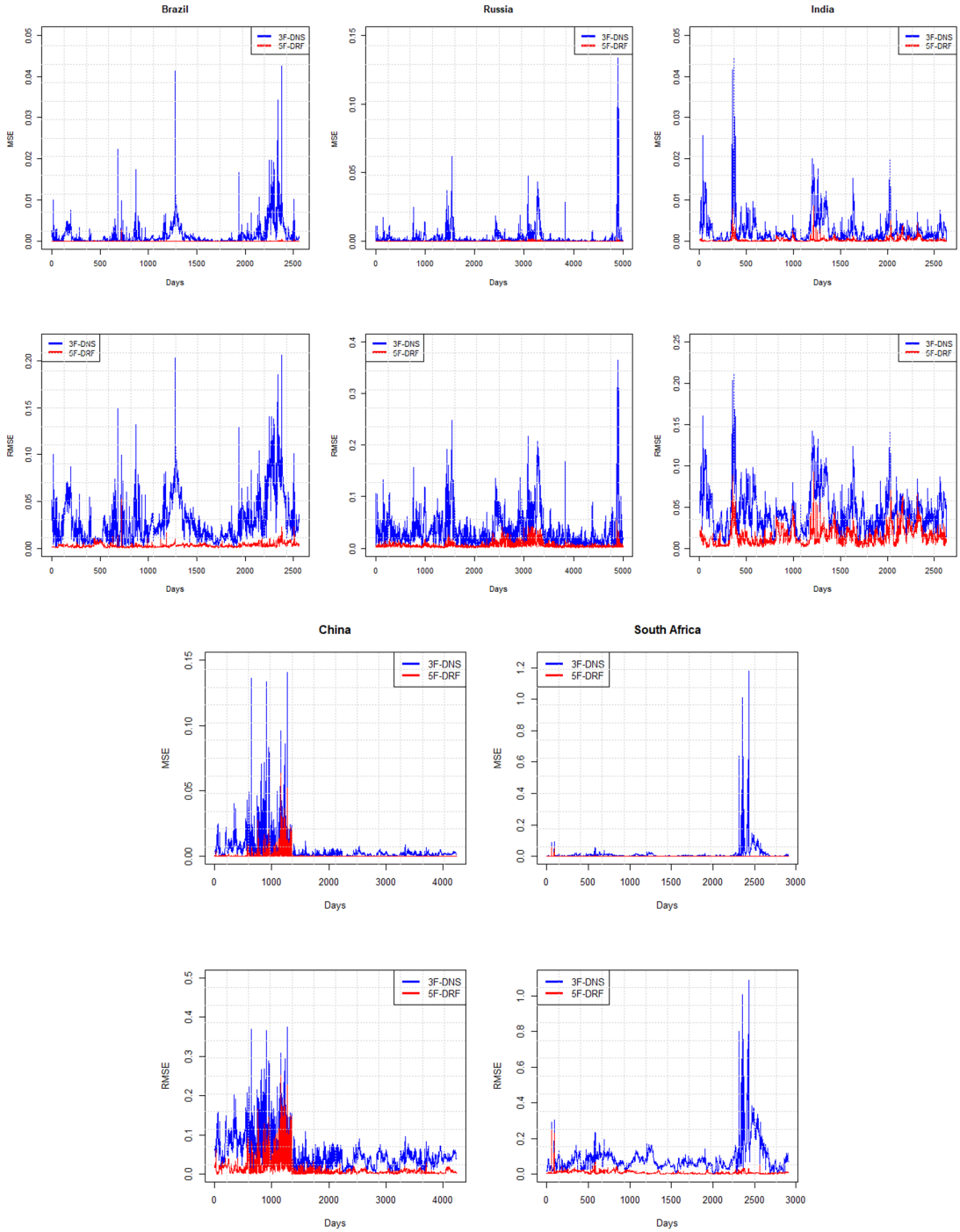


Figure 2.8 – *For each BRICS country (in column) the graph compares MSE (first line) and RMSE (second line) time series generated by the 3F-DNS and 5F-DRF models.*

Chapter 3

Modeling and Forecasting Natural Gas Futures Prices Dynamics: An Integrated Approach

3.1 Introduction

During recent decades, the liberalization and financialization wave (Cheng and Xiong, 2014, Creti and Nguyen, 2015) 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% (Futures Industry Association (FIA), 2022) 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 (Furtuna et al., 2022).

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 Pascual and Zambetakis (2016), 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 Bordoff and O'Sullivan (2023), 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 Bartuska et al. (2019) 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 (Snam, 2020). 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 (Chen et al., 2019, Guidolin and Alpcan, 2019).

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 (Schoenberg, 1946), four-factor dynamic Nelson–Siegel–Svensson (4F-DNSS) (Svensson, 1994), and five-factor dynamic De Rezende–Ferreira (5F-DRF) (De Rezende and Ferreira, 2008) 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 (Svensson, 1994, Lin, 2002, Koopman et al., 2007, De Pooter, 2007, Caldeira et al., 2010, De Rezende and Ferreira, 2013, Muthoni et al., 2015, Chouikh et al., 2017, Ullah and Bari, 2018, Faria and Almeida, 2018, Nunes et al., 2019, Nagy, 2020, Mineo et al., 2020). 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 (Guo and Xue, 2014, Baruník and Malinská, 2016, Ruiz et al., 2016, Zhou

et al., 2018, Aliberti et al., 2019, Benrhmach et al., 2020, Di Franco and Santurro, 2020, Butler et al., 2021, Chi, 2021, Xu and Zhang, 2021, Blanchard and Samanta, 2020, Xu and Zhang, 2022), 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, Emery and Liu, 2002, Brown and Yücel, 2008, Brigida, 2014, Etienne et al., 2016, Gatfaoui, 2016, Zhang et al., 2017, Li et al., 2017, Ji et al., 2018, Behmiri et al., 2019, Tiwari et al., 2019), as well as on modeling price volatility (e.g., Suenaga et al., 2008, Lv and Shan, 2013, van Goor and Scholtens, 2014, Saltik et al., 2016, Fałdziński et al., 2021, Lu et al., 2022, Liang et al., 2022, Guo et al., 2023), demand and supply (e.g., Szoplik, 2015, Khan, 2015, Shaikh and Ji, 2016, Panapakidis and Dagoumas, 2017, Shaikh et al., 2017, Chen et al., 2018, Ö zmen et al., 2018, Hribar et al., 2019, Su et al., 2019), spot prices (e.g., Mu, 2007, Panella et al., 2012, Salehnia et al., 2013, Mason and N.A. Wilmot, 2014, Jin and Kim, 2015, Čeperić et al., 2017, Dolatabadi et al., 2018, Berrisch and Ziel, 2020, Kwas and Rubaszek, 2021, Li et al., 2021, Wang et al., 2021, Pei et al., 2023) or futures prices of individual contracts (e.g., Borovkova and Mahakena, 2015, Jana and Ghosh, 2022, Li and Song, 2023). 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. (2009) proposed a two-factor regime-switching volatility model enhanced with the Markov chain Monte Carlo estimation method to model the forward price curve; Almansour (2016) studied futures curve dynamics with an extension of the Gibson and Schwartz (1990) two-factor model in a regime-switching framework; Leonhardt et al. (2017) 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. (2017) studied futures prices comovements of the most traded commodities with a factor approach relying on the Diebold and Li (2006) model; Jablonowski and Schicks (2017) introduced a three-factor model based on Heath et al. (1992) to describe the relationship between gas term structure and temperature forecasts. Finally, Tang et al. (2019) 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 (2019) investigated the abilities of GARCH-type discrete-time models and different Poisson jump-diffusion models to fit NG futures data, while Horváth et al. (2020) 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 bring 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 3.2 analyzes the features and main stylized facts of the data set; Section 3.3 presents the methodologies in use for modeling and forecasting, respectively; Section 3.4 discusses the main results; Section 3.5 concludes the paper.

3.2 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 (Heather, 2021), 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 (Intercontinental Exchange, 2021).

Figure 3.1 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.

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.

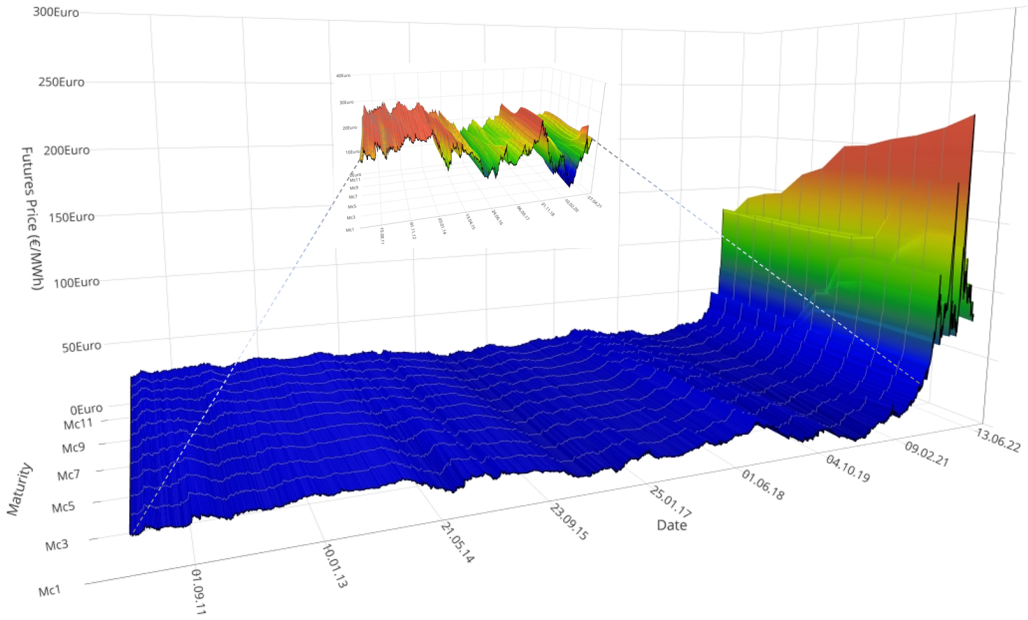


Figure 3.1 *– 3D surface plot of the term structure of natural gas futures prices: the x-axis shows the time expressed in days, the y-axis represents the maturities from 1 month (Mc1) to 12 months (Mc12), and the z-axis the price of the contracts in Euros. The data spans 2916 trading days, from 20 January 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.*

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 (Copernicus Climate Change Service (C3S), 2022); (ii) low levels of gas storage, with underground storage facilities less than 77% full throughout 2021 and less than 57% in December 2021 (Gas Infrastructure Europe (GIE), 2022); (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 (2006). We, therefore, argue that stylized facts in the bonds and in the NG futures market are similar. To examine these properties, in Table 3.1 we present the main descriptive statistics of futures prices and of the daily volatility (σ_{daily}) of futures prices series for each NG futures contract calculated as the absolute value of price returns, following Bessembinder et al. (1996).

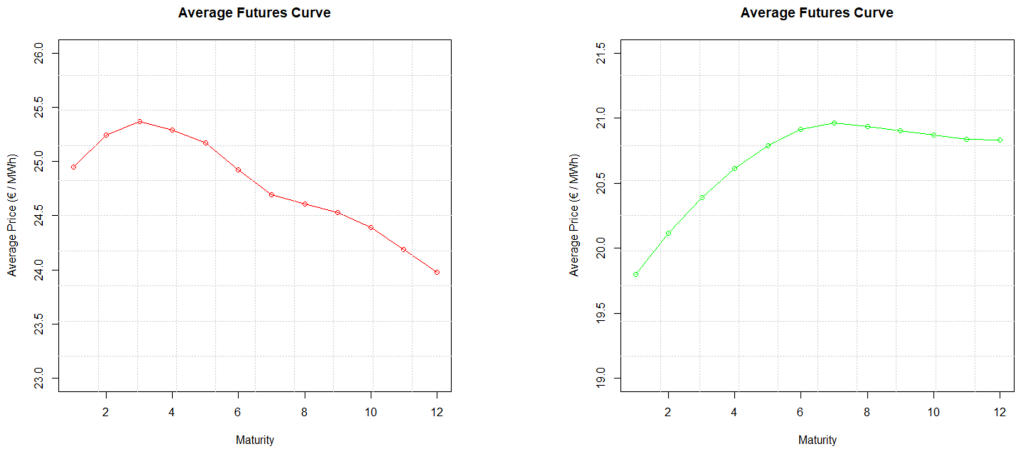
Table 3.1 – *Descriptive statistics of prices and daily volatility for each natural gas futures contract. For the price, we reported the mean, the standard deviation (SD), the minimum (Min), and the maximum (Max) values, while for the daily volatility (σ_{daily}) we examined the mean and the median.*

	Price				σ_{daily}	
Maturity	**Mean**	**SD**	**Min**	**Max**	**Mean**	**Median**
***Mc*1**	24.951	20.928	3.509	227.201	2.031	1.167
***Mc*2**	25.244	20.764	4.058	217.293	1.870	1.088
***Mc*3**	25.369	20.170	4.618	210.804	1.743	1.000
***Mc*4**	25.288	19.095	5.406	206.905	1.688	0.942
***Mc*5**	25.170	18.327	7.082	200.902	1.611	0.923
***Mc*6**	24.923	17.376	7.921	199.052	1.557	0.913
***Mc*7**	24.690	16.645	9.194	179.233	1.507	0.906
***Mc*8**	24.605	16.410	10.692	171.752	1.449	0.894
***Mc*9**	24.527	16.078	11.130	154.291	1.405	0.848
***Mc*10**	24.394	15.544	10.828	149.990	1.368	0.806
***Mc*11**	24.190	14.733	10.801	143.515	1.329	0.797
***Mc*12**	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 3.2 (**a**) 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 3.2 (**b**)),

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 3.2 (**a**), 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.



(a) ***(b)***

Figure 3.2 *– 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 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 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 3.1 (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 (Samuelson, 1973), 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 Duong and Kalev (2008), Jaeck and Lautier (2014) and ran the Jonckheere–Terpstra test (JT test) (Jonckheere, 1954, Terpstra, 1952) 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 σ_{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 3.2, indicate that the σ_{daily} time series are not normally distributed, they are autocorrelated, and they do not contain unit roots.

Table 3.2 *– 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 (σ_{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.*

	σ_{daily}		
Maturity	**JB Test**	**LB Test**	**ADF Test**
***Mc*1**	1.87×10^5 *	4.07×10^3 *	-9.628 *
***Mc*2**	2.35×10^5 *	4.23×10^3 *	-9.376 *
***Mc*3**	3.65×10^5 *	3.98×10^3 *	-9.131 *
***Mc*4**	8.67×10^5 *	3.53×10^3 *	-8.068 *
***Mc*5**	7.79×10^5 *	3.91×10^3 *	-8.485 *
***Mc*6**	8.12×10^5 *	4.21×10^3 *	-8.728 *
***Mc*7**	6.19×10^5 *	4.48×10^3 *	-8.144 *
***Mc*8**	7.13×10^5 *	4.78×10^3 *	-7.911 *
***Mc*9**	7.75×10^5 *	5.24×10^3 *	-8.204 *
***Mc*10**	7.59×10^5 *	5.11×10^3 *	-8.088 *
***Mc*11**	8.23×10^5 *	4.92×10^3 *	-7.952 *
***Mc*12**	1.49×10^6 *	4.51×10^3 *	-7.238 *

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:

$$\begin{cases} H_0 : \tilde{\sigma}_{12} = \tilde{\sigma}_{11} = \ldots = \tilde{\sigma}_1 \\ H_1 : \tilde{\sigma}_{12} \leq \tilde{\sigma}_{11} \leq \ldots \leq \tilde{\sigma}_1 \end{cases} \tag{3.1}$$

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 2.59×10^8 and a p–value of 2×10^{-16}, allow to reject the H_0 hypothesis at 1% significance level, thus supporting 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 3.3 shows five *slices* extracted from the 3D surface plot, representing the main trends observed on the NG market in different periods.

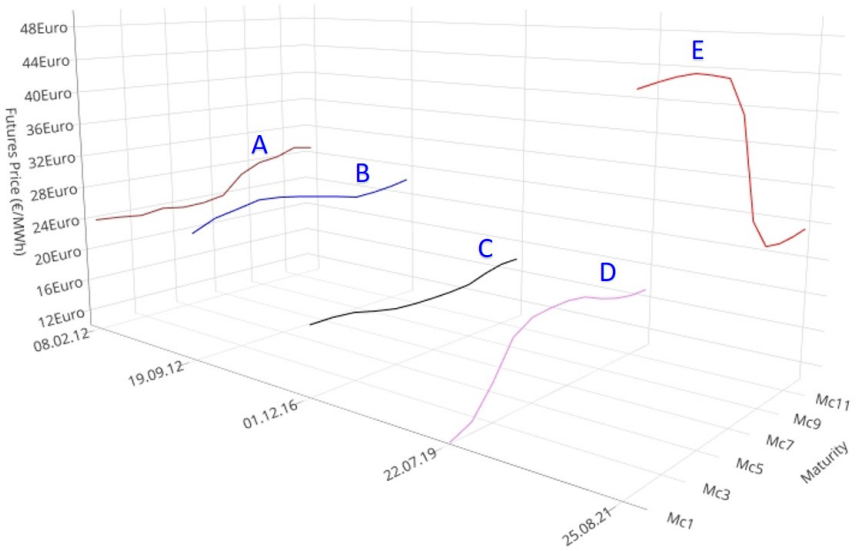


Figure 3.3 *– Plot of the main futures curves shapes observed on the market at various times t = 8 February 2012, 19 September 2012, 1 December 2016, 22 July 2019, and 25 August 2021. Time is represented on the x-axis, while maturities and prices (€/MWh) are on the y-axis and z-axis, respectively.*

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.

3.3 Modeling Approach

3.3.1 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 (Nelson and Siegel, 1987) 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 $\boldsymbol{p}(t)$ the $N \times 1$ vector of observed gas futures prices available at maturity $m \in M = (m_1, m_2, \ldots, m_N)'$, where N is the maximum maturity length in months, and consider a time horizon of length t, $t = 1, \ldots, T$ expressed in days. The price dynamic is described by:

$$\boldsymbol{p}(t) = \boldsymbol{F}(t)\boldsymbol{\beta} + \boldsymbol{\eta}(t). \tag{3.2}$$

The variables in (3.2) deserve some further explanation. We start with $\boldsymbol{F}$, 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:

$$F_m^{DNSS}(t) = \left[1 \quad \tau_1 \tfrac{1-e^{-m/\tau_1}}{m} \quad \tau_1 \tfrac{1-e^{-m/\tau_1}}{m} - e^{-m/\tau_1} \quad \tau_2 \tfrac{1-e^{-m/\tau_2}}{m} - e^{-m/\tau_2}\right], \tag{3.3}$$

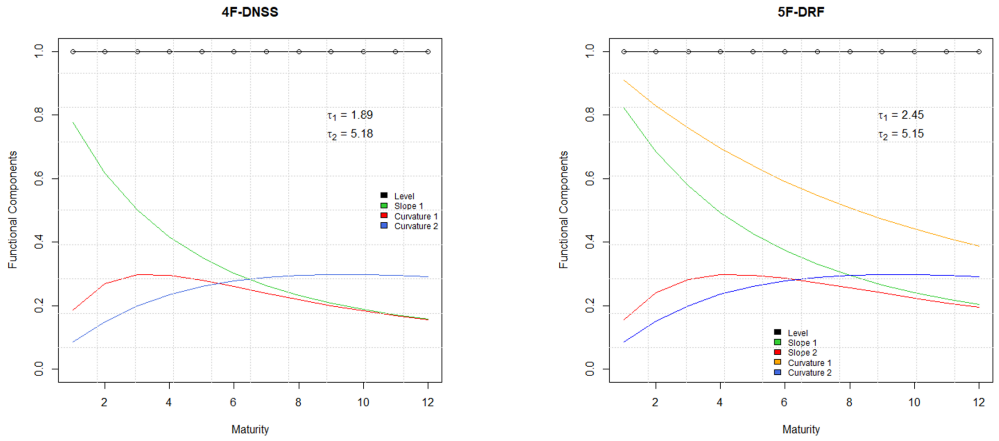
in the 4F-DNSS model, i.e., when $k = 4$, or:

$$F_m^{DRF}(t) = \left[1 \quad \tau_1 \tfrac{1-e^{-m/\tau_1}}{m} \quad \tau_2 \tfrac{1-e^{-m/\tau_2}}{m} \quad \tau_1 \tfrac{1-e^{-m/\tau_1}}{m} - e^{-m/\tau_1} \quad \tau_2 \tfrac{1-e^{-m/\tau_2}}{m} - e^{-m/\tau_2}\right], \tag{3.4}$$

in the 5F-DRF model, i.e., when $k = 5$; here, τ_1 and τ_2 are the decay terms which regulate the exponential components' decaying speed.

Three elements characterize the $\boldsymbol{F}$ matrix, that is, the factor loadings, i.e., the building blocks of futures curves. The first element, and also the first component in both (3.3) and (3.4) is the level that represents the long-term component, constant for every maturity. The second element, occupying position 2 in (3.3) and positions 2 and 3 in (3.4) 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 (3.3) and in positions 4 and 5 in (3.4), 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 (3.3) and (3.4) 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 3.4 shows the behavior of the factor loadings in the case of the 4F-DNSS (a) and 5F-DRF (b) models.



(a) *(b)*

Figure 3.4 *– 4F–DNSS* ***(a)*** *and 5F–DRF* ***(b)*** *factor loadings at different time to maturity. In* ***(a)*** *and* ***(b)*** *we indicate the average values of* τ_1 *and* τ_2 *determined during the estimation process and used for the daily fits.*

Turning to $\boldsymbol{\beta}$, we have $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3]'$ in the 4F–DNSS model, and $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_4]'$ in the 5F–DRF model. Each element represents the weight associated to the corresponding factor loading, hence, changes in the vector $\boldsymbol{\beta}$ 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, $\boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ represents the error terms vector, assumed normally distributed with zero mean vector and variance-covariance matrix $\boldsymbol{\Sigma}$.

For what is concerning the estimation process of $\boldsymbol{\beta}$, we applied an approach organized into two stages. At first, following Castello and Resta (2022), we identified for each time t the optimal combination $[\hat{\tau}_1(t), \hat{\tau}_2(t)]$, and hence $\hat{\boldsymbol{\beta}}(t)$, 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 $\hat{\boldsymbol{\beta}}^*(t)$ for each available

day. In this way the model maintains a high adaptive capability and gains in stably estimated parameters.

3.3.2 B–Spline interpolation method

B–Spline (Schoenberg, 1946, Curry and Schoenberg, 1947, 1966) is a powerful modeling tool to fit observable data without strong functional form assumptions. The B–Spline function is:

$$f(x) = \sum_{i=1}^{k+d-2} \pi_i B_{i,d}(x), \tag{3.5}$$

where π_i $(i = 1, \ldots, k+d-2)$ are the spline coefficients and $B_{i,d}(x)$, $d \geq 1$ are B–Spline basis functions. Those, in turn, are fully determined once set the order $d \geq 1$ and the sequence of nondecreasing real values $\xi_1 \leq \xi_2 \leq \ldots \leq \xi_k$, acting as control points or knots. To have 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:

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

The i–th B–Spline basis of order d is then recursively defined, for $d > 1$, as:

$$B_{i,d}(x) = \delta_{i,d}(x) B_{i,d-1}(x) + [1 - \delta_{i+1,d}(x)] B_{i+1,d-1}(x), \tag{3.6}$$

with:

$$B_{i,1}(x) = \begin{cases} 1, & \text{if } \xi_i \leq x \leq \xi_{i+1} \\ 0, & \text{otherwise} \end{cases}, \tag{3.7}$$

and:

$$\delta_{i,d}(x) = \begin{cases} \dfrac{x - \xi_i}{\xi_{i+d-1} - \xi_i}, & \xi_i \neq \xi_{i+d-1} \\ 0, & \text{otherwise} \end{cases}. \tag{3.8}$$

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 π_i: in general, the closer the consecutive π_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 3.5 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 $\mathscr{C}^1$) and curvature (curve segments with same second derivative at joint, i.e., functions belonging to $\mathscr{C}^2$) were applied at each knot, except to those overlapping; here, the B-spline is of class $\mathscr{C}^0$, that is, curve segments are connected at the joint.

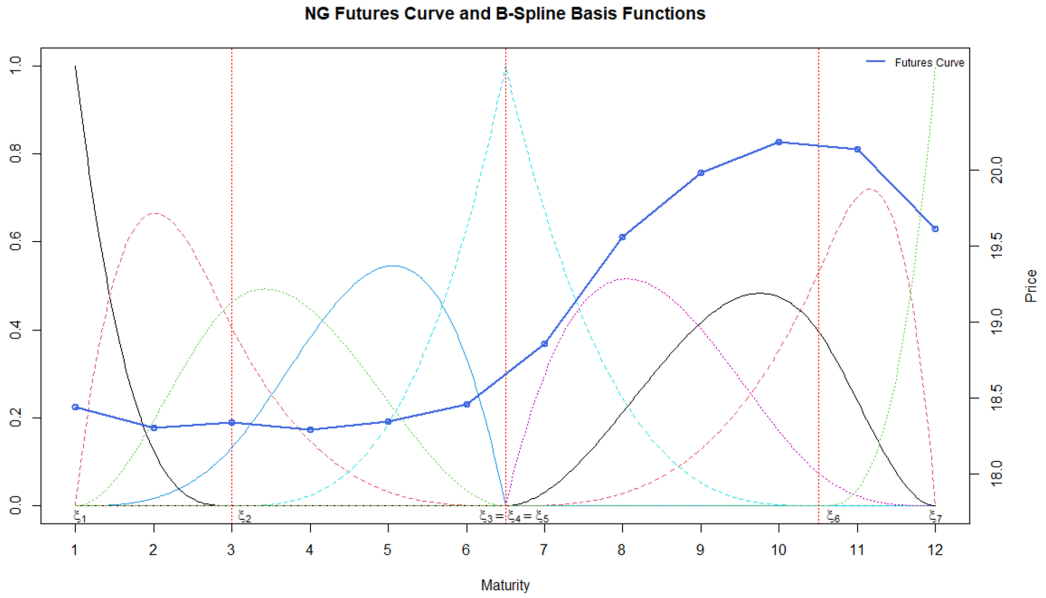


Figure 3.5 – *The plot shows an example of the NG futures curve (blue line), the corresponding* 4*–order basis polynomial functions (dashed lines), and the* $\xi_1, \ldots, \xi_7$ *knots.*

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

$$WSSE(\boldsymbol{\pi}(t)) = \sum_{j=1}^{N} \omega(m_j, t)[p(m_j, t) - f(m_j, t)]^2. \tag{3.9}$$

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.

3.3.3 Nonlinear Autoregressive Neural Network (NAR–NN)

An Artificial Neural Network is a system aimed at simulating the human nervous system. It is characterised by a computational scheme which is not programmed but trained by a machine learning algorithm. Thanks to the ability in identifying nonlinear relationships in the data, it can approximate any differentiable function (Reed and Marks, 1999) and it is recognized as a universal approximator (Hornik et al., 1989).

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 $m \in M = (m_1, m_2, \ldots, m_N)'$ at time t, and h past values or feedback delays for each $t = 1+h, \ldots, T$, capturing the autoregressive properties:

$$x_{m,t} = g(x_{m,t-1}, x_{m,t-2}, \cdots, x_{m,t-h}) + \epsilon_{m,t}, \tag{3.10}$$

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 (Rosenblatt, 1958), 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 $x_{m,t-j}$, $j = 1, \ldots, h$. Each input pattern is used as the input to the network. Its elements are multiplied by an assigned weight $\theta_{i,j}$ $(j = 1, \ldots, h)$ 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:

$$\varphi_{i,t} = \theta_{i,0} + \sum_{j=1}^{h} \theta_{i,j}\ x_{m,t-j}. \tag{3.11}$$

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 χ, 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 $\hat{x}_{m,t}$:

$$\hat{x}_{m,t} = \psi\left(\gamma_0 + \sum_{i=1}^{n} \gamma_i \Lambda_\chi\left(\varphi_{i,t}\right)\right), \tag{3.12}$$

where γ_i is the weight assigned to the connection between the hidden unit i and the output unit; and γ_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 $\boldsymbol{\nu}^*$ of weights and bias by means of a learning algorithm that minimizes the error function:

$$Err(\boldsymbol{\nu}) = \sum_{t=1+h}^{T} (x_{m,t} - \hat{x}_{m,t})^2, \tag{3.13}$$

After the training phase, the neural network is converted into a closed loop network and, for $t > T$, the $(x_{m,t-1}, x_{m,t-2}, \ldots, x_{m,t-h+1}, x_{m,t-h})$ 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 $(\hat{x}_{m,t}, x_{m,t-1}, x_{m,t-2}, \ldots, x_{m,t-h+1})$, 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:

$$\Lambda_{\chi}(\varphi_{i,t}) = \frac{1}{1+e^{-\varphi_{i,t}}}, \tag{3.14}$$

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%) set. The supervised learning process was carried out implementing the Levenberg–Marquardt Back Propagation learning algorithm (LMBP) (Levenberg, 1944, Marquardt, 1963b) with the weights update rule:

$$\boldsymbol{\Delta\nu}_k = -\left(\boldsymbol{J}^T(\nu_{k-1})\boldsymbol{J}(\nu_{k-1}) + \mu\boldsymbol{I}\right)^{-1}\boldsymbol{J}^T(\nu_{k-1})\ \boldsymbol{e}_{k-1}. \tag{3.15}$$

where $\boldsymbol{J}$ is the Jacobian matrix of the network errors with respect to the weights and biases; μ represents the damping factor; $\boldsymbol{I}$ is the Identity matrix; while $\boldsymbol{e_{k-1}}$ 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 μ and update weights moving in the steepest–descent direction. If the update fails to reduce the error, then μ 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.

3.4 Empirical Study

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

3.4.1 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 3.3 lists average descriptive statistics of the fitted prices for each available maturity. The interpretation is twofold: a first reading key concerns the models performances, while another relates to the models ability to replicate stylized facts.

By comparison with the descriptive statistics summarized in Table 3.1, 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 3.6 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 3.4 shows the performance metrics computed on the models' residuals generated during the estimation process, while Figure 3.7 presents a 2D visualization of the mean absolute error (MAE) metric generated by the 4F-DNSS, 5F-DRF, and B-spline models.

Table 3.3 *– 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 RMSE at every maturity.*

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

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.

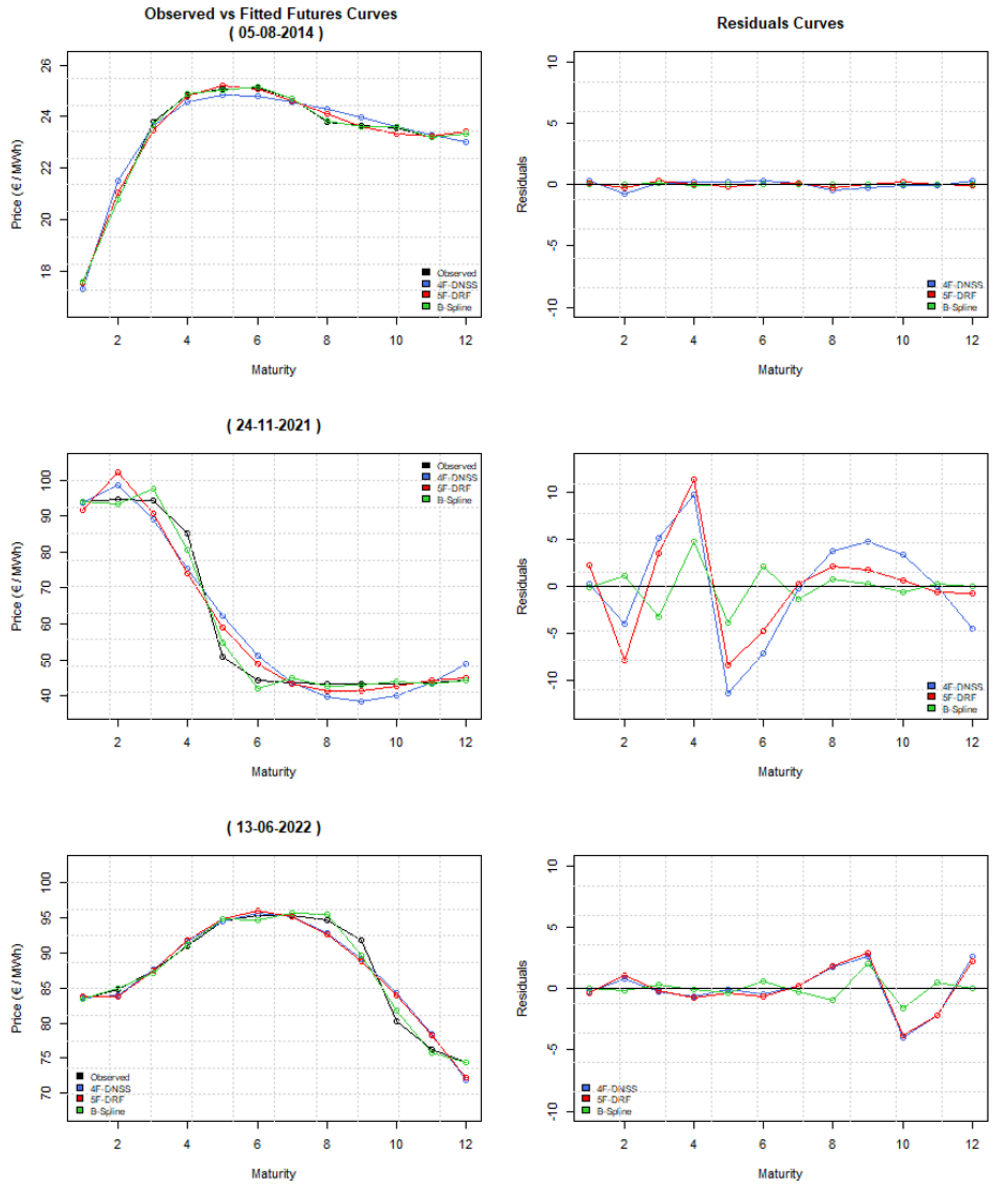


Figure 3.6 *– The first column shows the observed and fitted futures curves with the 4F-DNSS (blue), 5F-DRF (red), and B-spline (green) models; the related residuals curves are shown on the right-hand side. The days chosen are representative of the most difficult dynamics of the futures curves.*

Table 3.4 *– Main MSE and RMSE statistics for the 4F-DNSS, 5F-DRF and B–Spline models. For each indicator 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**
Mean	1.3586	8.0050×10^{-1}	1.3978×10^{-1}	5.8370×10^{-1}	3.9148×10^{-1}	1.4050×10^{-1}
SD	6.5667	4.5393	9.5190×10^{-1}	1.0091	8.0460×10^{-1}	3.4653×10^{-1}
Min	3.9908×10^{-3}	3.2177×10^{-4}	2.5177×10^{-5}	6.3173×10^{-2}	1.7938×10^{-2}	5.0177×10^{-3}
Max	98.9464	83.6844	17.4452	9.9472	9.14792	4.17675

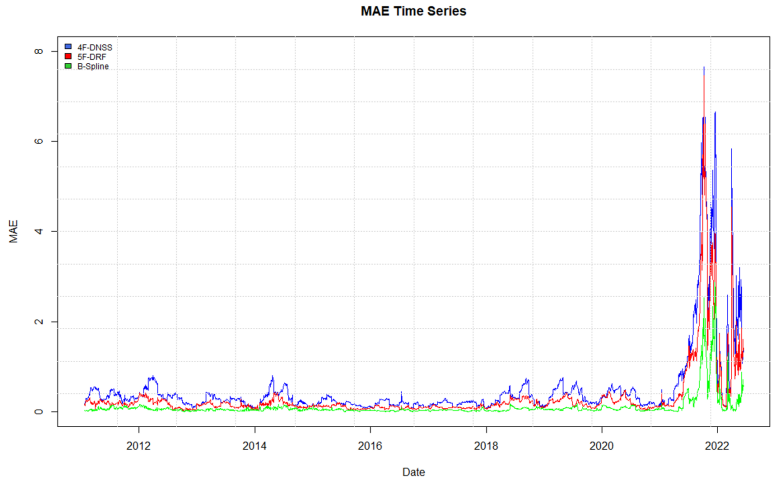


Figure 3.7 *– MAE time series obtained with the 4F-DNSS (blue), 5F-DRF (red) and B-Spline (green) models in the Natural Gas futures market.*

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.

3.4.2 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 (2006) 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 forecasting process was organized into four main stages. First, we collected time series of the models estimated parameters and time series of futures prices, each to be used as input to the neural network. Then, we set up the NAR–NNs based on the features of each collected time series. Next, using the expanding window approach, we performed one–day–ahead forecasts of the parameters as well as of the prices time series for 21 working days in 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. We hence derived the forecasted NG futures curves from the predicted models parameters. Finally, we evaluated the effectiveness of the proposed approach exploiting the the Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{T}\sum_{t=1}^{T}\frac{1}{M}\sum_{m=1}^{M}\left|\frac{p_{t+h,m}-\hat{p}_{t+h,m}}{p_{t+h,m}}\right|, \tag{3.16}$$

and the Mean Squared Forecast Error (MSFE)

$$MSFE = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{M}\sum_{m=1}^{M}(p_{t+h,m}-\hat{p}_{t+h,m})^2, \tag{3.17}$$

as performance metrics and we reported them in Table 3.5.

Table 3.5 *– Average MAPE (%) and MSFE to test models forecasting accuracy.*

	4F–DNSS/NAR–NN	**5F–DRF/NAR–NN**	**B–Spline/NAR–NN**	**NAR–NN**
MAPE	6.1309	7.6939	2.7562	2.7619
MSFE	43.4550	85.4639	9.2254	9.4914

Based on Table 3.5 all the candidate methods achieved satisfying results, ensuring a prediction accuracy ranging on average between 92.3 % and 97.2 %. Looking at the results, however, it is possible to create a rank on models performance with the 5F–DRF/NAR-NN at the 4-th place with the highest MAPE, the 4F–DNSS/NAR–NN at the 3-rd place, the NAR–NNs gains the second place, while the top position is taken by the B–Spline/NAR–NN.

A better understanding of the results can be gained with an example: Figure 3.8 **(a)** and **(b)** compare futures curves observed on June 7 and 13, 2022 with those predicted according to the proposed methodology; 3.8 **(c)** and **(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 enough close to the observed ones for each available maturity, without notable spikes or outliers.

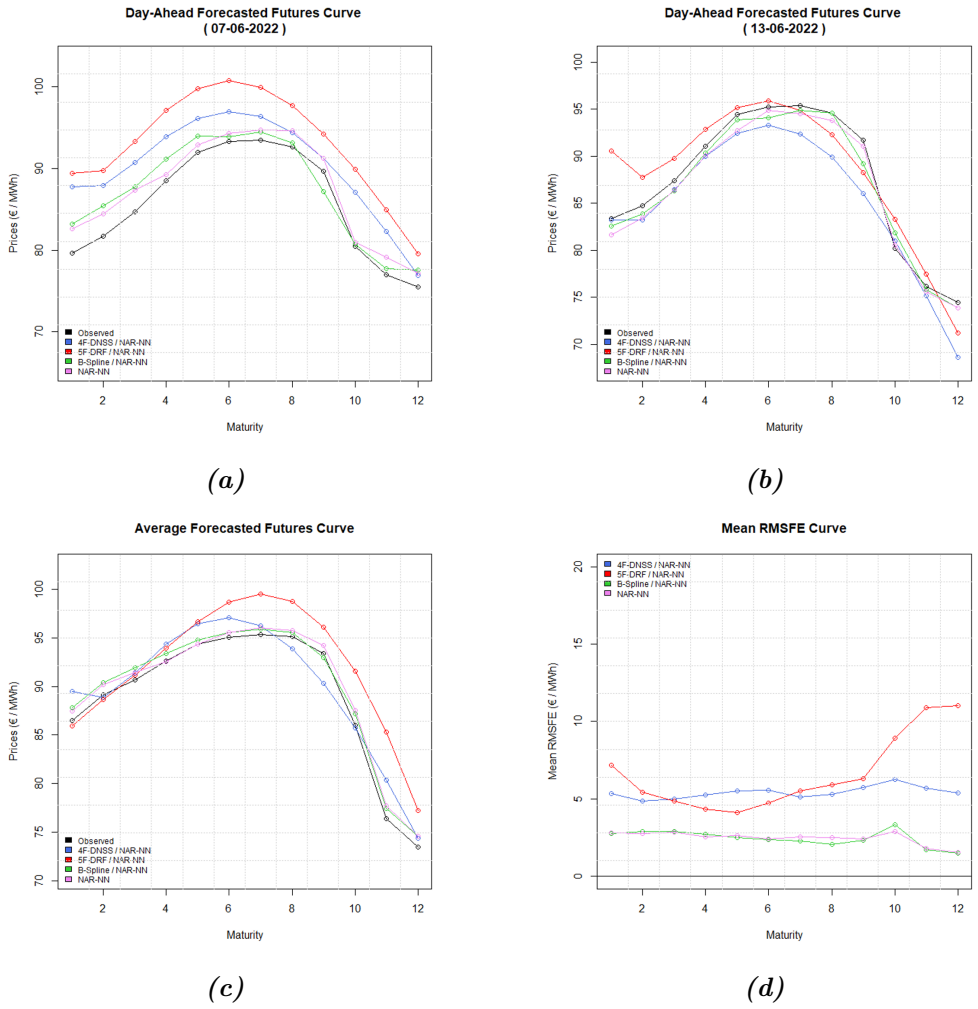


Figure 3.8 *– Comparison of day-ahead forecasts in some worst cases (**a**,**b**). Average predictions (**c**) and average RMSFE (**d**) generated by the 4F-DNSS (blue), 5F-DRF (red), B-spline (green), and NAR-NNs (violet) models.*

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.

Overall, the reason for the weaker performances of both the factor models can be probably found in the volatility of the $\boldsymbol{\beta}$ 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.

3.5 Conclusion

In this chapter 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.

Chapter 4

Univariate and Multivariate Electricity Futures Curve Forecasting using Artificial Neural Network Models

4.1 Introduction

Electricity is an essential source of energy in modern societies, playing a pivotal role in powering various segments of the economy such as the residential, commercial, industrial, and transport sectors. Therefore, changes in its prices and availability can exert deep, beneficial – as well as potentially devastating – knock-on effects on the social and economic fabric, affecting the development prospects of countries.

Exogenous events such as climate change, green economy policies, deregulation, as well as the geopolitical crises of recent times, have exerted a significant impact on world energy markets, particularly the electricity market, increasing uncertainty and causing huge price volatility with sudden spikes and drops in price levels (da Silva and Horta, 2019, Ghosh et al., 2021).

The primary reason for the sensitivity of the price of electricity to exogenous events lies in specific features that distinguish it from any other energy commodity, namely, the characteristics of "non–storability" and inelasticity of demand. These facts make it complex to manage the influence on prices of sudden changes in supply and demand due to exogenous shocks, consequently exposing market participants to significant risk in power prices.

In light of this, risk management plays a key role in mitigating the negative consequences deriving from excessive exposure to market price volatility, and derivative instruments such as futures are one of the most widespread financial instruments used to achieve this objective.

Electricity futures traded on financial markets have different maturities, and all together form the so called Term Structure of Futures Prices. The term structure describes the relationship between the settlement prices of futures contracts and different expiration dates, with the futures curve being its graphical representation. The analysis of its evolution over time gives an idea of the markets supply and demand dynamics, as well as practitioners expectations about the markets behaviour and the potential development direction.

Setting aside speculative purposes, electricity futures contracts are used by the counterparties with the principal objective of reducing their risk exposure. Power producers, in particular, will take short positions on futures contracts to lock in a guaranteed price, thereby hedging against any downward movements in the spot price and decreasing the volatility of their revenues. Consumers and power retailers, for their part, will take long positions, taking advantage of the different electricity delivery periods in contracts (from a few days to weeks, a season, or a whole year), to buy/sell the good at a fixed price, thereby reducing the risk of encountering sudden price spikes.

All this leads to greater stability in price and supply, with direct benefits on both the social fabric and the economy in terms of development and competitiveness. In this respect obtaining reliable forecasts of the futures curves dynamics and trends becomes of paramount importance in many scenarios and for different market players.

To date, there is a significant amount of research studies around electricity spot price modeling and forecasting. Extensive reviews of the state–of–the–art have been provided in Weron (2014), Fanelli et al. (2016), Nowotarski and Weron (2018), Yang et al. (2020), Deschatre et al. (2021), Lago et al. (2021), Olivares et al. (2022). According to Weron (2014) the most widely used approaches use statistical/econometric methods (Conejo et al., 2005, Tan et al., 2010, Voronin et al., 2013, Raviv et al., 2015, Zhao et al., 2017, Ziel and Weron, 2018, Kumar et al., 2018), Machine Learning (ML) (Peng et al., 2018, Zhou et al., 2019, Jahangir et al., 2020, Yang and Schell, 2021, Yang et al., 2022, Olivares et al., 2022, Tschora et al., 2022, Su et al., 2022), probabilistic and hybrid techniques (Dong et al., 2011, Gonzalez et al., 2012a, Wan et al., 2014, Dudek, 2016, Rafiei et al., 2017, Kuo and Huang, 2018, Ziel and Steinert, 2018, Kostrzewski and Kostrzewska, 2019, Zhang et al., 2019, Cheng et al., 2019, Zhang et al., 2020b,a, de Marcos et al., 2020, Kavitha and Kalpana, 2023), as well as fundamental (Eydeland

and Wolyniec, 2002, Karakatsani and Bunn, 2008, Coulon and Howison, 2009, Gonzalez et al., 2012b, Dillig et al., 2016, Ribeiro et al., 2020, Maciejowska et al., 2021, Kanamura and Bunn, 2022) or reduced–form models (Higgs and Worthington, 2008, Borovkova and Schmeck, 2017, Pircalabu and Benth, 2017, Xiong and Mamon, 2019, Apergis et al., 2019, Deng et al., 2020, Oduor, 2022).

However, research around predicting futures prices dynamics is relatively more limited and less diversified. Most of the literature focuses mainly on deriving futures prices from spot prices with various pricing techniques based on stochastic processes, rather than on direct price prediction. To cite main research stream structural methods (Füss et al., 2015, Kallabis et al., 2016) model electricity price dynamics considering the influence of fundamental drivers (e.g. demand, loads, weather conditions); while reduced–form methods (Koekebakker and Ollmar, 2005, Benth and Koekebakker, 2008, Itoh and Kobayashi, 2010, Barth and Benth, 2014, Islyaev and Date, 2015, Biagini et al., 2015, Fanelli et al., 2016, Gudkov and Ignatieva, 2021, Mehrdoust and Noorani, 2021, Najafi et al., 2023), explain price dynamics based on their main statistical features. Relatively to ML techniques, to date their predictive abilities within electricity futures markets haven't been extensively analyzed like for spot markets. To the best of the authors knowledge, except Yu et al. (2008), Zhang and Shi (2009) and Nascimento et al. (2019), there are no other published papers exploring the potentials of computational intelligence methods to forecast directly electricity futures prices and/or the entire term structure.

Attempting to contribute to this research strand, this study develops a framework based on Dynamic Recurrent Neural Network (RNN) models to tackle the issue of predicting the Electricity Futures Prices Term Structure dynamics. In particular we make use and evaluate the Nonlinear Autoregressive Neural Networks (NAR–NNs), Nonlinear Autoregressive Neural Network with Exogenous Inputs (NARX–NNs), Long Short–Term Memory Neural Networks (LSTM–NNs) and Encoder–Decoder Long Short–Term Memory Neural Networks (ED–LSTM–NNs) for day–ahead predictions of the electricity futures curve. Furthermore, aimed at verifying whether incorporating relevant exogenous influence factors can lead to improvements in the models forecasting power, a setting based on the multivariate forecasting approach is also explored. The purpose of our work is therefore twofold: we asses the potentials of the models to produce robust electricity futures curves predictions using, on the one hand side, only the autoregressive dynamics of the target series that form the futures curve; while on the other we consider the latter information jointly with the information content contained in the more closely related covariates.

The main contribution of this work to the existing literature can be summarized as follows:

(i) we conducted day–ahead forecasts of electricity futures prices term structure by means of Dynamic Recurrent Neural Networks;

(ii) albeit some ANN models have been used in previous researches, to the best of our knowledge this study is the first to utilize NAR–NNs, NARX–NNs, LSTM–NNs and ED–LSTM–NNs neural networks using both the univariate and multivariate prediction methods.

(iii) we extend the idea behind fundamental models to neural networks by incorporating different time–dependent exogenous input features in order to enhance neural networks predictive abilities.

(iv) to stress–test the models performances, we collected daily data from March 2017 to September 2022 and tested the proposed framework over both stable and extremely volatile periods.

The remainder of the paper is organized as follows. Section 4.2 discusses the RNN models in use; Section 4.3 analyzes the data set; Section 4.4 presents the empirical findings and discusses the main results; Section 4.5 concludes.

4.2 Recurrent Neural Networks–Based Models

Recurrent Neural Networks (RNNs) are a class of deep neural networks conceived for processing sequential or time series data for predictions. They key features of RNNs are the loop–based architecture and memory mechanisms which enables the network to store past information, learn sequences and/or long–term dependencies of the data, and utilize this knowledge to improve the performance of the network on current and future outputs of the sequence.

In this section we present and discuss three types of RNNs that have gained wide success for univariate and multivariate forecasting of time–series. In particular we focus our attention on Nonlinear–Autoregressive (NAR–NNs), Nonlinear–Autoregressive with Exogenous Inputs (NARX–NNs), Long Short–Term Memory (LSTM–NNs) as well as Encoder–Decoder Long Short–Term Memory Neural Networks (EDLSTM–NNs).

4.2.1 Nonlinear Autoregressive Neural Networks

The Nonlinear Autoregressive Neural Network (NAR–NN) and the Nonlinear Autoregressive Neural Network with Exogenous variables (NARX–NN) belong to the class of Recurrent Neural Networks employed for one and multi-step ahead forecasts of time series data. They are conceived as a nonlinear generalization of the classical linear Autoregressive (AR) model, aimed at managing nonlinear processes.

Given the univariate time series $\{x_t\}_{t=1}^{T}$, the NAR–NN predicts the future value at time t based exclusively on its d past endogenous values (feedback delays) used as regressors:

$$x_t = f(x_{t-1}, x_{t-2}, \cdots, x_{t-d}), \tag{4.1}$$

where $f(\cdot)$ represents an unknown nonlinear transfer function that the network tries to approximate while x_{t-k} $(k = 1, \ldots, d)$ are the endogenous lagged input signals.

With regard to the NARX–NN model, it represents a modified version of the above NAR–NN and it allows to take advantage from the inclusion of information from exogenous inputs. The NARX–NN predicts the future value of the univariate series $\{x_t\}_{t=1}^{T}$ at time t using both the d past endogenous values of the x_t series and the h lagged values of the exogenous series $\{y_t\}_{t=1}^{N}$:

$$x_t = g(x_{t-1}, x_{t-2}, \cdots, x_{t-d}, y_{t-1}, y_{t-2}, \cdots, y_{t-h}), \tag{4.2}$$

where x_{t-k} $(k = 1, \ldots, d)$ is the same as above, $g(\cdot)$ is an unknown nonlinear transfer function, while y_{t-j} $(j = 1, \ldots, h)$ represents the exogenous lagged input observations.

The NAR–NN and NARX–NN models are characterized by a fully–connected Multilayer Feed–Forward Network architecture in which interconnected information processing elements (neurons) are grouped in multiple sequential layers, namely the Input Layer which propagates the input features to the adjacent Hidden Layer; one or more Hidden Layers representing the computational step between the Input and Output layers where the network tries to learn and recognize complex patterns in the data by means of nonlinear transformations of the inputs; and the final Output Layer which leads to the predicted value $\hat{x}_t$. Additionally the networks contains a re–feeding mechanism which allows the use of predicted values as inputs for future forecasts. The structure of the NAR–NN and NARX–NN models is demonstrated in Figure 4.1.

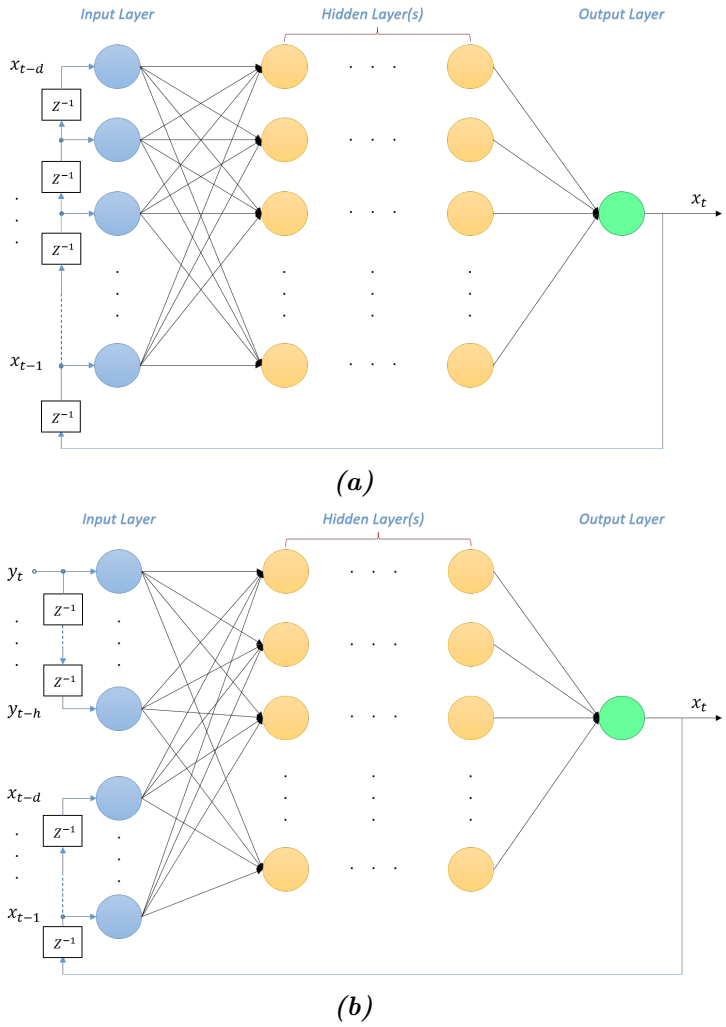


Figure 4.1 *– Topology of the NAR–NN (a) and NARX–NN (b) models.*

The optimal topology, i.e. the number of hidden layers, neurons, feedback delays, activation functions, depends on the problem domain and it is usually carried out by means of trial and error approach by comparing different network configurations, taking also into consideration the bias/variance tradeoff. Once set, the network undergoes the training and learning process in an open loop where lagged input series are made up only of observed data. In this phase, adopting the Supervised Learning strategy networks parameters are calibrated by means of the back–propagation method combined with an optimization algorithm, such that the overall networks error is minimized. Different learning algorithms can be used to optimize the loss function, the choice depends on factors like computational time, error goals, or the amount of training data, weights and biases. The Levenberg–Marquardt Backpropagation Algorithm (LMBP) (Levenberg, 1944, Marquardt, 1963a) is one of the most widely used learning rules in NAR–NNs

and NARX–NNs as it inherits the advantages of the steepest descent method (in terms of stability) and of the Gauss–Newton algorithm (in terms of speed), hence ensuring robustness, fast convergence rate and quicker training speed (Wilamowski and Irwin, 2011). In the end, the configuration that provides the best performance is chosen. After the training the network is turned into a Closed Loop network and is used to execute one/multistep–ahead predictions. During this process the predicted values are fed back towards the input layer to serve as new input signals for successive predictions.

From a mathematical viewpoint, the relationship between x_t and the endogenous/exogenous feedback delays that replaces the function $f(\cdot)$ is given by

$$x_{k,t+h} = \omega_{k,0} + \sum_{i=1}^{n} \omega_{k,i} \; \Lambda \left(\alpha_{i,0} + \sum_{j=1}^{d} \alpha_{i,j} x_{k,t+h-j} \right) + \varepsilon_{k,t} \tag{4.3}$$

for the NAR–NN, or

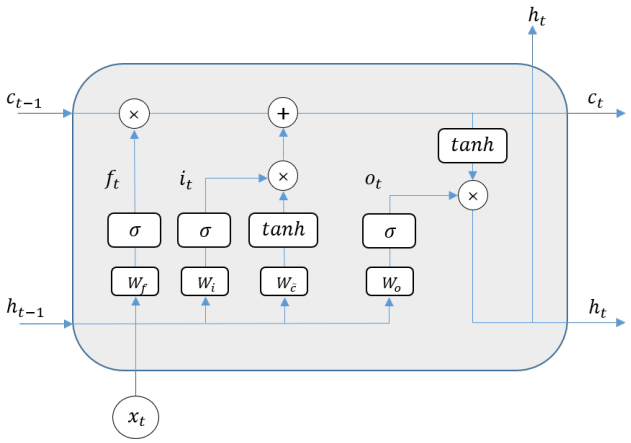
$$x_{k,t+h} = \omega_{k,0} + \sum_{i=1}^{n} \omega_{k,i} \; \Lambda \left(\alpha_{i,0} + \sum_{j=1}^{d} \left(\alpha_{i,j} x_{k,t+h-j} + \gamma_{i,j} y_{k,t+h-j} \right) \right) + \varepsilon_{k,t} \tag{4.4}$$

for the NARX–NN; here, in either model d is the number of lagged endogenous ($x_{k,t+h-j}$) and/or exogenous ($y_{k,t+h-j}$) input units fed to each hidden neuron i; Λ is an activation function that applies a specific transformation to the weighted sum of inputs and biases, and is responsible for the activation or the deactivation of the networks neurons; $\alpha_{i,j}$ and $\gamma_{i,j}$ represent the neuron weight of the connection between the input unit j and the hidden unit i; $\omega_{k,i}$ is the weight of the connection between the hidden unit i and the output unit k; $\alpha_{i,0}$ and $\omega_{k,0}$ are the bias used to optimize the working point of the neurons in the hidden and output units respectively; finally $\varepsilon_{k,t}$ represents the error term.

4.2.2 Long Short–Term Memory Network

The Long Short–Term Memory Neural Network (LSTM–NN) is a type of RNN developed by Hochreiter and Schmidhuber (1997). It is capable of learning both short and long–term nonlinear dependencies in sequential data and retain relevant information over a long–term period, thus overcoming the RNNs problem of vanishing/exploding gradients which hampers their ability to learn long data sequences.

The LSTM–NN is characterized by a chain–like architecture with a linked sequence of repeating modules known as cells (or memory units): Figure 4.2 presents the structure of a LSTM–NN cell.



(a)

Figure 4.2 *– Schematic representation of the LSTM memory unit.*

The cell has a gating mechanism that allow the creation of long temporal relationships. Each cell contains three types of gates, namely the *Forget gate*, the *Input gate* and the *Output gate* which direct/regulate information flow within the LSTM unit deciding which information should be remembered and which instead ignored.

In detail, the *Forget gate* has the task to determine whether the information from the previous time step has to be remembered or can be ignored. In particular, the gate receives the hidden state h_{t-1} and the current input value x_t as inputs and process them via the logistic function which returns values between 0 and 1. If the result is 0 then the information is completely discarded, otherwise it is retained. The output of the *Forget gate* is described by

$$f_t = \sigma\left(W_f h_{t-1} + W_f x_t + b_f\right), \tag{4.5}$$

$$\sigma(\theta) = \frac{1}{1+e^{-\theta}}, \tag{4.6}$$

where $\sigma(\theta)$ is the logistic activation function, while W_f and b_f represent the weight matrix and connection bias, respectively.

On the other hand, the *Input gate* determines which new information must be kept and added to update the current cell state. The gate receives the hidden state h_{t-1} and the current value x_t and process them via both the logistic and tanh functions.

The latter generates a vector of new candidate values while the sigmoid output decides which information will be retained from the *tanh* output. The new information to update the current cell state is obtained combining the result of the two functions with the information retained from the previous cell state (c_{t-1}). The computation steps are explained on following

$$i_t = \sigma\left(W_i h_{t-1} + W_i x_t + b_i\right), \tag{4.7}$$

$$\tilde{c}_t = tanh\left(W_{\tilde{c}} h_{t-1} + W_{\tilde{c}} x_t + b_{\tilde{c}}\right), \tag{4.8}$$

$$c_t = f_t \odot c_{t-1} + i \odot \tilde{c}_t, \tag{4.9}$$

$$tanh(\theta) = \frac{e^{\theta} - e^{-\theta}}{e^{\theta} + e^{-\theta}}, \tag{4.10}$$

Here i_t represents the input threshold at time t, $\tilde{c}_t$ is the temporary cell state at time t, $tanh(\cdot)$ is the hyperbolic tangent activation function, c_t is the final cell state at time t, $\odot$ is the Hadamard product, while W_i, $W_{\tilde{c}}$, b_i and $b_{\tilde{c}}$ are as before.

Finally, the *Output gate* calculates the value of the new hidden state h_t, i.e. the cell output. This is achieved firstly by passing the values of the current input and previous hidden state through the sigmoid function. Then the obtained output is combined with the output of the tanh function thus determining which part of the information encoded in the updated cell state will be used as output:

$$o_t = \sigma\left(W_o h_{t-1} + W_o x_t + b_o\right), \tag{4.11}$$

$$h_t = o_t \odot tanh(c_t), \tag{4.12}$$

where o_t is the output value of the gate at time t, W_o and b_o are as before while h_t is the LSTM cell output signal at time t.

4.2.3 Encoder–Decoder LSTM model

The Encoder–Decoder LSTM neural network (Sutskever et al., 2014) represents a combinatorial architecture based on the principles of Encoder–Decoder and LSTM networks used for solving the problem of mapping sequences to sequences characterized by variable dimensionality (Brownlee, 2018).

The network consists of three main components: the encoder and the decoder subnetworks and an intermediate fixed–length context vector. Internally, the encoder and decoder are made by one or multiple layers of LSTM cells representing their structural blocks. The topology of the Encoder–Decoder LSTM neural network is shown in Figure 4.3.

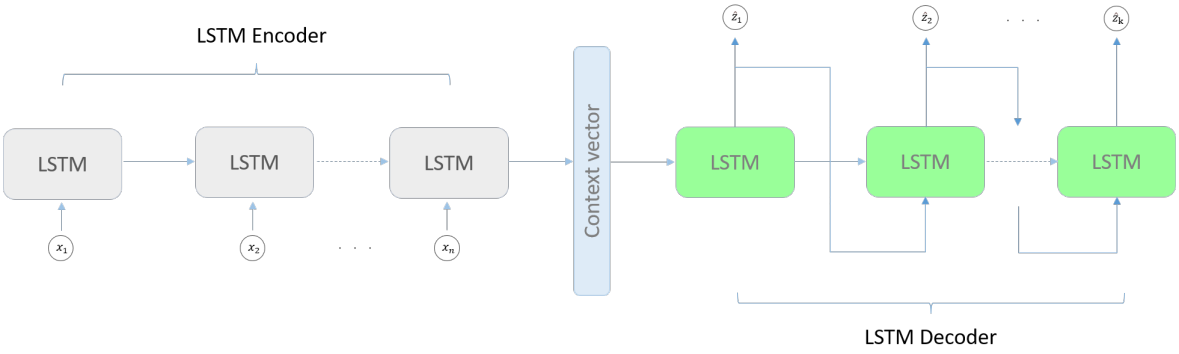


Figure 4.3 – *Encoder–Decoder LSTM Architecture.*

The encoder operating principle is based on the LSTM network method of learning the long–term dependencies described in Subsection 4.2.2. Given a variable–length vector $x = (x_1, \ldots, x_n)$ as the input sequence to the model, each LSTM unit in the encoder sequentially processes the incoming data, retains relevant features, and propagates the information forward through the sequence chain. In the final step, the encoder squeezes and encodes the information into a fixed–length context vector, that is, the final cell ($\boldsymbol{c}$) and hidden state ($\boldsymbol{h}$) vectors, which in turn become the input to the decoder. At this stage, the decoder is initialised with the ($\boldsymbol{c}$) and ($\boldsymbol{h}$) vectors jointly with the last input value and starts the recursive generation of the output sequence $\hat{z} = (\hat{z}_1, \ldots, \hat{z}_k)$ one time step at a time, using the obtained output as input for successive updates.

4.3 Data

Our dataset consists of daily settlement prices of electricity futures contracts. The contracts are quoted in €/MWh and traded on the European Energy Exchange (EEX), which represents Europe's leading marketplace for energy and commodity products. We considered contracts with seven different maturities: from 1 up to 6 month representing the short to medium term contracts and we labeled them Ec1–Ec6; furthermore we also considered 1 year to expiration, interpreted as the long–term contract, and we labeled it by Ec12. The total number of observation values is 1399 covering the period

from March 3, 2017 up to September 15, 2022. Such a wide time–span ensures a considerable amount of data containing different stable and spiky periods with futures curves exhibiting frequent temporary reversals. This enables us to verify and validate the effectiveness and robustness of the proposed framework under different market conditions.

In Appendix, Figure 4.8 illustrates the evolution of the prices time series of the seven futures contracts together with their first differences. In between January 2021 and September 2022 it is possible to observe an unprecedented increase of both price levels and volatility across all the given maturities mainly driven by the recent global COVID–19 pandemic and the new Russia–West confrontation (Khudoley, 2022) that caused acute global supply chain and energy shortage issues. Table 4.1 provides the descriptive statistics of the electricity futures prices dataset and their first differences, and include the following measures: Mean, Standard Deviation (SD), Median, Minimum (Min), Maximum (Max), Skewness and Kurtosis.

Table 4.1 *– Descriptive statistics of futures prices and their logarithmic transformation for each maturity.*

	Price					**Log Price**						
Maturity	**Mean**	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**	**Skew**	**Kurtosis**
Ec1	99.36	104.75	55.52	21.39	581.90	1.7×10^{-3}	3.3×10^{-2}	3.8×10^{-4}	-0.26	0.33	1.77	30.79
Ec2	102.41	110.56	56.56	25.25	738.65	1.8×10^{-3}	3.7×10^{-2}	5.9×10^{-4}	-0.30	0.33	0.27	17.54
Ec3	103.70	111.59	58.50	27.72	793.75	1.9×10^{-3}	3.8×10^{-2}	6.9×10^{-4}	-0.32	0.36	8.2×10^{-3}	20.01
Ec4	103.54	111.91	59.02	32.06	804.47	1.8×10^{-3}	3.9×10^{-2}	1.2×10^{-3}	-0.32	0.33	-0.32	20.96
Ec5	102.18	108.91	59.70	35.35	801.14	1.7×10^{-3}	4.0×10^{-2}	9.0×10^{-4}	-0.47	0.33	-1.21	29.46
Ec6	100.28	104.89	60.93	35.69	732.41	1.7×10^{-3}	3.6×10^{-2}	1.2×10^{-3}	-0.33	0.28	-0.25	17.05
Ec12	86.88	78.43	58.78	41.25	624.58	1.6×10^{-3}	2.8×10^{-2}	1.6×10^{-3}	-0.59	0.19	-7.19	154.84

The data confirm high volatility in each price series; huge positive spikes can be observed for all maturities in the period of greater markets turmoil, with record high peaks going over 700 €/MWh on average; by comparing the latter outcome with the average values calculated using the Min e Median columns of the price series of Table 4.1, it turns out that it is 23 times larger than the lowest price level and 12 times larger than the average median value.

Moving to the returns series and analyzing their distribution, results show significant departure from Normality. The *Ec1, Ec2, Ec3* series are characterized by positive skewness; on the contrary, results for *Ec4, Ec5, Ec6, Ec12* indicate a leptokurtic left–skewed distribution with greater probability of negative outlying returns.

Furthermore, we have collected the data of multiple influence assets which relevantly affect electricity futures price dynamics. For this task we have chosen the Natural Gas (NG) and Coal futures prices, as well as the spot price series of the Carbon Emission Certificates (CO_2). These variables were treated as time–dependent exogenous features and used in conjunction with the electricity futures prices as inputs to neural networks within the multivariate forecasting process. The datasets of the exogenous variables were sourced from Thomson Reuters Datastream. In Appendix, Figure 4.9 shows the temporal evolution of electricity prices at each maturity along with the related exogenous variables. To ensure uniformity with the data structure of electricity futures, all the considered exogenous series are made up of daily values covering the same time horizon, from March 3, 2017 up to September 15, 2022. Moreover, for the NG and Coal futures contracts we selected the same maturities as well which we labeled as NGc1, NGc2, NGc3, NGc4, NGc5, NGc6, NGc12 and as Cc1, Cc2, Cc3, Cc4, Cc5, Cc6, Cc12 respectively; relatively to the CO_2 spot contract we labeled it as COc1.

The choice of these variables is motivated by the relevant degree of correlation with electricity futures prices. This means that highly likely they do embed useful information about the movements of electricity prices and represent at the same time its driving forces. For measure the strength of the relationship occurring between the considered datasets, we calculated the Pearson correlation coefficient (ρ) between each electricity futures price series with the related counterparty in the NG, Coal and CO_2 datasets, illustrating the obtained results in Table 4.2

Table 4.2 *– Correlation between each electricity futures price series and the related price series of Natural Gas (first column), Coal (second column) and CO_2 (last column) contracts.*

Maturity	ρ_{ElecNG}	$\rho_{ElecCoal}$	ρ_{ElecCO_2}
Ec1	0.35	0.30	0.014
Ec2	0.74	0.50	0.14
Ec3	0.75	0.49	0.15
Ec4	0.70	0.49	0.14
Ec5	0.71	0.48	0.11
Ec6	0.69	0.43	0.18
Ec12	0.64	0.34	0.22

It can be seen that electricity futures are significantly affected by all the influence factors; the strongest positive correlation is shared with the NG data, with an average value of over 0.65.

4.4 Empirical Evaluation

4.4.1 Experiment Setup

The purpose of this study is to explore the capabilities of the RNN models introduced in Section 4.2 to make stable and robust predictions of electricity futures curves. The model comparison is performed on the basis of a forecasting process carried out following two alternative settings. Figure 4.4 shows the designed forecasting process synthesized in the graphic form of a flowchart.

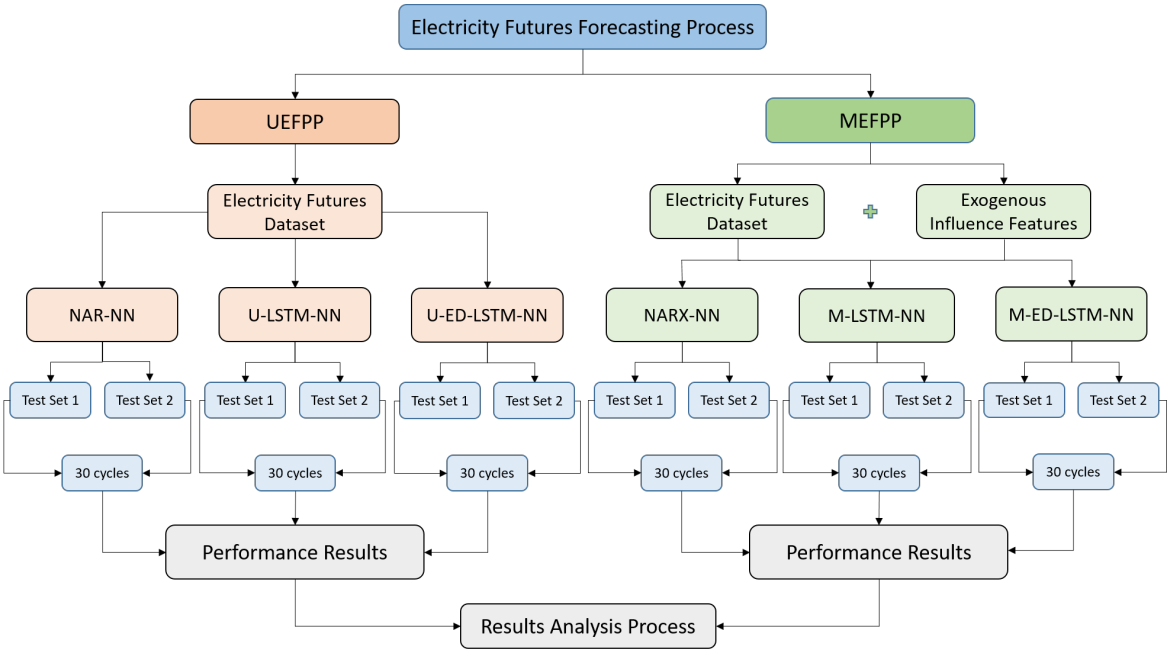


Figure 4.4 – *Workflow of the proposed forecasting framework.*

As a first approach, we used the univariate electricity futures price prediction process (UEFPP). For this task the NAR–NN, the univariate LSTM–NN (U–LSTM–NN) and the Univariate Encoder–Decoder LSTM–NN (U–EDLSTM–NN) models were implemented to predict prices at each maturity based only on its past data readings, that is using only the information content of its own past values. From a mathematical viewpoint this is given by

$$Ec_{i,t} = f(Ec_{i,t-1}, Ec_{i,t-2}, \ldots, Ec_{i,t-k}), \qquad i = 1,2,3,4,5,6,12. \tag{4.13}$$

where $Ec_{i,t}$ is the predicted price value, while $Ec_{i,t-k}$ are the k past values of the i-th maturity representing the networks input vector.

Relatively to the second approach, futures curves prediction was carried out using the multivariate prediction process (MEFPP): we used the NARX–NN, Multivariate LSTM (M–LSTM–NN) and Multivariate Encoder–Decoder LSTM (M–EDLSTM–NN) methods to predict prices of each maturity individually, based on both its historical data and that of the exogenous covariates described in Section 4.3. The MEFPP strategy can be mathematically expressed as

$$Ec_{i,t} = f(Ec_{i,t-k}, NGc_{i,t-k}, Cc_{i,t-k}, COc_{1,t-k}), \qquad i = 1,2,3,4,5,6,12 \tag{4.14}$$

where $NGc_{i,t-k}$, $Cc_{i,t-k}$ and $COc_{1,t-k}$ represent the k past values of the Natural Gas and Coal futures prices time series at maturity i and CO_2 spot price series respectively.

Moreover, given the unprecedented turmoil on energy markets observed during last years, we decided to divide the whole dataset into two subperiods and conduct both the UEFPP and MEFPP on two distinct test sets to validate their predictive power. The first subperiod, with the related Test Set 1, was chosen as the more stable and less volatile one, while the second subperiod, with the related Test Set 2, as the one characterized by price spikes and overall higher variance.

Test Set 1 ranges from September 16, 2019 to July 8, 2020 while Test Set 2 covers the period from November 22, 2021 to September 9, 2022; each set is made up of 204 daily futures prices. Tables 4.3 and 4.4 provide descriptive statistics of the two test sets. It is possible to observe that the level of prices across the spectrum of maturities and the standard deviation of Test Set 2 are, at least, 6 and 10 times higher compared to Test Set 1, respectively.

Table 4.3 *– Descriptive statistics of electricity futures prices for Test Set 1.*

Maturity	**Ec1**	**Ec2**	**Ec3**	**Ec4**	**Ec5**	**Ec6**	**Ec12**
Mean	38.98	41.93	44.49	46.51	47.54	47.92	51.81
SD	10.85	10.92	10.83	10.46	8.59	5.92	4.99
Min	21.39	25.25	27.72	32.06	35.35	35.69	42.96
Max	56.75	61.00	63.53	66.35	68.55	65.82	62.12

Table 4.4 *– Descriptive statistics of electricity futures prices for Test Set 2.*

Maturity	**Ec1**	**Ec2**	**Ec3**	**Ec4**	**Ec5**	**Ec6**	**Ec12**
Mean	317.78	329.09	330.20	326.84	317.66	312.33	243.68
SD	109.94	126.26	132.64	142.28	143.76	134.68	106.15
Min	201.49	189.07	182.00	162.37	128.94	121.92	122.20
Max	581.90	738.65	793.75	804.47	801.14	732.41	624.58

For either test set in both the UEFPP and MEFPP settings, each maturities price prediction was conducted on a daily basis using the sliding window method, for an overall number of 408 predicted days.

As pointed out in Kim et al. (2020) there is no rule of thumb for networks topology and hyperparameter optimization. We therefore carried out a trial and error approach and examined various models settings considering different combinations for feedback delays, activation functions, number of nodes, cells, hidden layers and learning algorithms. Relatively to the NAR–NN and NARX–NN methods a good compromise was found selecting one hidden layer, 5 delays, and 10 hidden neurons for NAR–NN and 15 for NARX–NN; additionally, for both models the sigmoid (linear) activation function was used for the hidden (output) neurons while the Levenberg–Marquardt Backpropagation learning rule was chosen for networks training and learning. With regard to the LSTM–NN and ED–LSTM–NN in both the univariate and multivariate settings, the best solution turned out to be: 1 layer with 200 units for the LSTM–NN and 1 layer with 100 units in both the encoder and decoder; both models were trained in a supervised learning fashion for 1000 epochs using the Adaptive Moment Estimation (ADAM) optimization algorithm (Kingma and Ba, 2015). The code for the implementation of the proposed methodology and models was developed in MATLAB R2022a for what concerns the NAR–NN and NARX–NN methods, and in Python (3.10.10) using the open–source Keras (2.4.3) library (Chollet et al., 2015) with TensorFlow (2.4.0) as backend for the LSTM–NN and ED–LSTM-NN models.

Additionally, given the stochastic nature of ANNs models, we run thirty repeated forecasts (cycles) and considered the mean of the forecasted results as the final models result. This approach allows to ensure robust results and reduce the randomness of the forecasts.

4.4.2 Results discussion

To comprehensively evaluate the predictive accuracy of the proposed models, this paper employs two widely used performance indicators, namely the Root Mean Squared Forecasting Error (RMSFE) and the Mean Absolute Percentage Error (MAPE). In Table 4.5 we report the evaluation metrics implemented in this research along with their formulas, where y_{t+h} and $\hat{y}_{t+h}$ represent the observed and predicted values respectively, n denotes the sample size.

Table 4.5 *– Forecasting Performance metrics.*

Metric	**Definition**	**Equation**
RMSFE	Mean Squared Forecasting Error	$\text{RMSFE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_{t+h} - \hat{y}_{t+h})^2}$
MAPE	Mean Absolute Percentage Error	$\text{MAPE} = \frac{100}{n}\sum_{j=1}^{n}\left\lvert\frac{y_{t+h}-\hat{y}_{t+h}}{y_{t+h}}\right\rvert$

The out–of–sample forecasting results of the implemented framework are firstly presented in Table 4.6, highlighting in bold the best performing method. For each test set and model we summarized the Mean, the Standard Deviation (SD), the Minimum and the Maximum of the RMSFE and MAPE accuracy metrics.

An analysis of the obtained results indicates that all the implemented techniques exhibit a highly satisfactory performance, ensuring an average MAPE score greater than 95% within Test Set 1 and greater than 91% within Test Set 2, thereby acting as good predictive techniques.

Considering Test Set 1, it is possible to establish that the NAR–NN model significantly outperforms the other competing methods reporting the best RMSFE and MAPE scores, followed by the U–ED–LSTM–NN, NARX–NN, and M–LSTM–NN models, which exhibit quite close results. More specifically, considering, for instance, the MAPE metric the NAR–NN, U–ED–LSTM–NN, NARX–NN, and M–LSTM–NN models achieved 98.04%, 97.77%, 97.62%, and 97.36% predictive accuracies, respectively. Moreover, the M–ED–LSTM–NN and U–LSTM–NN models reported the lowest accuracy measures of 96.51% and 95.51%, respectively.

In this context, it can be noted that the incorporation of exogenous influencing factors provides a relevant boost only to the LSTM–NN performance. In fact, the M–LSTM–NN specification reduced the RMSFE score by 32%.

Table 4.6 *– Average MSFE and MAPE (%) metrics obtained with different competing methods.*

		RMSFE				**MAPE** (%)			
	Method	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*
Test Set 1	NAR–NN	**1.24**	**0.23**	**0.74**	**1.41**	**1.96**	**0.47**	**1.08**	**2.55**
	U–LSTM–NN	2.18	0.39	1.33	2.43	4.49	1.08	2.26	5.56
	U–ED–LSTM–NN	1.31	0.32	0.71	1.75	2.23	0.67	1.14	3.18
	NARX–NN	1.43	0.31	0.74	1.67	2.38	0.62	1.16	3.00
	M–LSTM–NN	1.49	0.28	0.99	1.92	2.64	0.71	1.59	3.87
	M–ED–LSTM–NN	1.92	0.94	1.20	3.84	3.49	2.02	1.82	7.51
Test Set 2	NAR–NN	48.12	16.20	20.12	66.74	8.01	2.19	3.50	10.09
	U–LSTM–NN	**23.86**	**4.87**	**16.77**	**28.37**	**4.95**	**0.94**	**3.51**	**5.80**
	U–ED–LSTM–NN	24.35	4.86	16.49	29.91	4.96	1.08	2.74	5.90
	NARX–NN	46.70	10.60	30.35	57.75	8.85	1.53	5.60	10.35
	M–LSTM–NN	25.66	3.40	20.74	29.47	5.33	0.84	3.48	5.87
	M–ED–LSTM–NN	31.93	10.26	17.59	47.18	7.07	2.82	3.25	11.25

With regard to Test Set 2, the U–LSTM–NN and U–ED–LSTM–NN models ensured the best performances (almost at par) in the day-ahead predictions, with 95.05% and 95.04% precision, respectively; furthermore, these models showed an almost halved RMSFE score with respect to the NAR–NN and NARX–NN models. Such results are achieved because the U–LSTM–NN and U–ED–LSTM–NN models enhance memory abilities which enables them to capture and store nonlinear patterns embedded in past futures price data. In the third place, we find the M–LSTM–NN model with 94.67% accuracy, followed by the M–ED–LSTM–NN model with 92.93% accuracy. Relatively to the NAR–NN model, which was the best in Test Set 1, it generated a poorer performance: the predictive precision decreased from 98.04% (Test Set 1) to 91.99% (Test Set 2), while the RMSFE increased by 38 times with respect to the score achieved in Test Set 1. A similar deterioration of the metrics is also observed for the NARX–NN model.

Notably, although the U–ED–LSTM–NN model outperformed the M–LSTM–NN model in terms of the MAPE, the latter is characterised by a lower variability of the results, indicating more stable predictions.

Another noteworthy aspect is that the addition of exogenous features does not bring meaningful improvements to the model families; nevertheless, the M–LSTM–NN

and M–ED–LSTM–NN models delivered better results in terms of both the RMSFE and MAPE indicators with respect to the NAR–NN and NARX–NN models.

Finally, it can be summarised and pointed out that the use of such test sets makes it possible to highlight the fact that LSTM–based models, in addition to performing well in stable market conditions, are also able to guarantee the maintenance of the initial performances even in extreme conditions, such as those characterising Test Set 2, without being particularly affected by turmoil. In fact, in the face of a significant increase in the price level across the maturity spectrum by at least 6 times, the reduction in predictive accuracy is maintained in the range of 0.46% – 3.58%, which remains above 93%. However, the NAR–NN model that performed better in first test set was unable to maintain its predictive power in the most unstable period.

The performance comparison of the univariate and multivariate models is furthermore shown in Figure 4.5 and 4.6 where we illustrate for each maturity in either test set, the models day–ahead forecasting results, and in Figure 4.7 where we display for each test set the average actual and predicted futures curves. Based on the visual analysis of the graphs, the following conclusions can be drawn: (i) all the models forecast trends are enough close to the observed trends and do not exhibit unreasonable spikes or drops; (ii) considering the price time series, NAR–NN and NARX–NN models tend to exhibit a more erratic behavior with respect to other methods; furthermore when peak values occur, the univariate and multivariate LSTM based models show a smoother overall behavior.

Overall, we can conclude that the proposed framework is able to achieve satisfying and stable predictive results within both stable as well as volatile market conditions, thus confirming the adequacy and usefulness of the implemented deep learning models for predicting electricity futures curves.

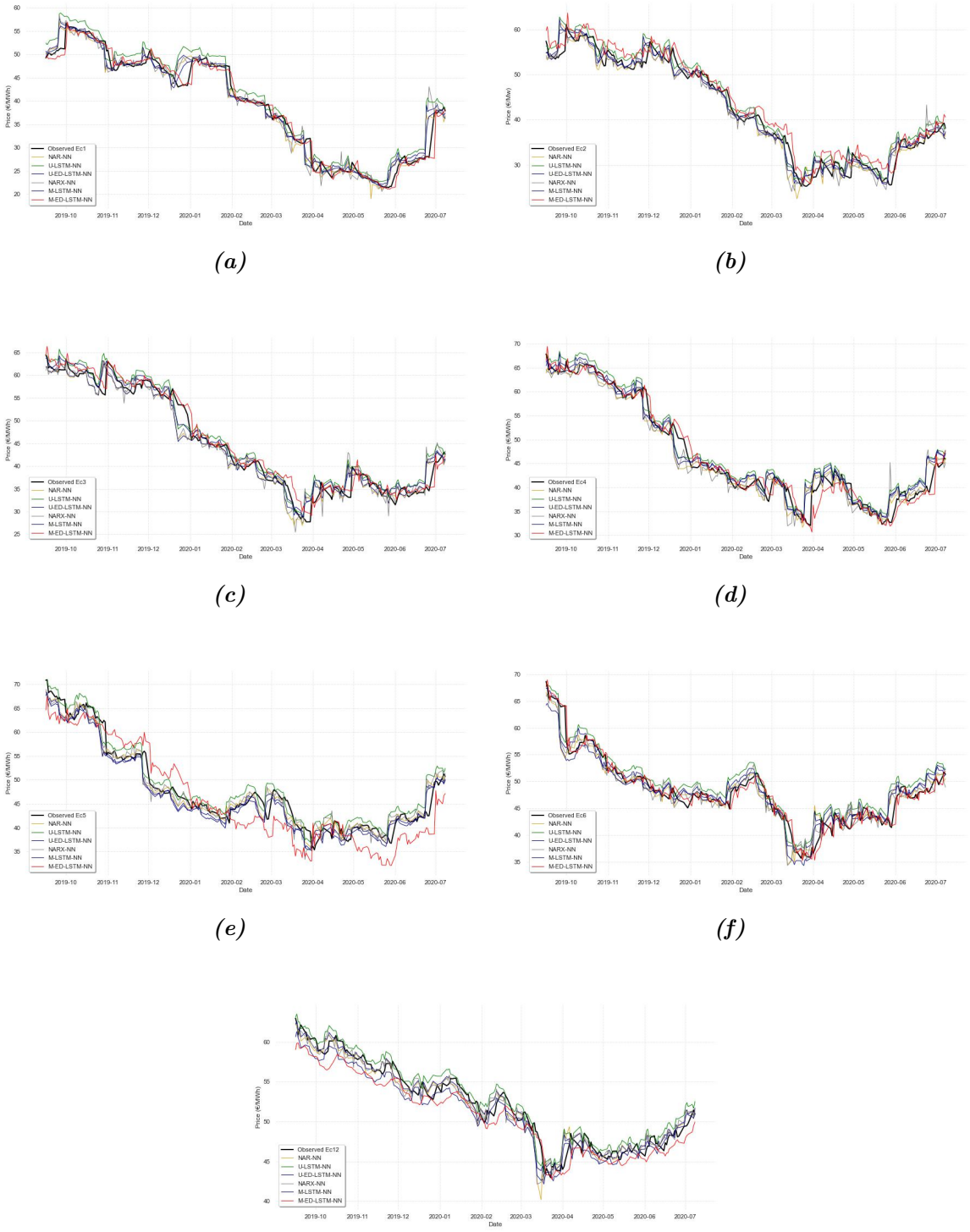
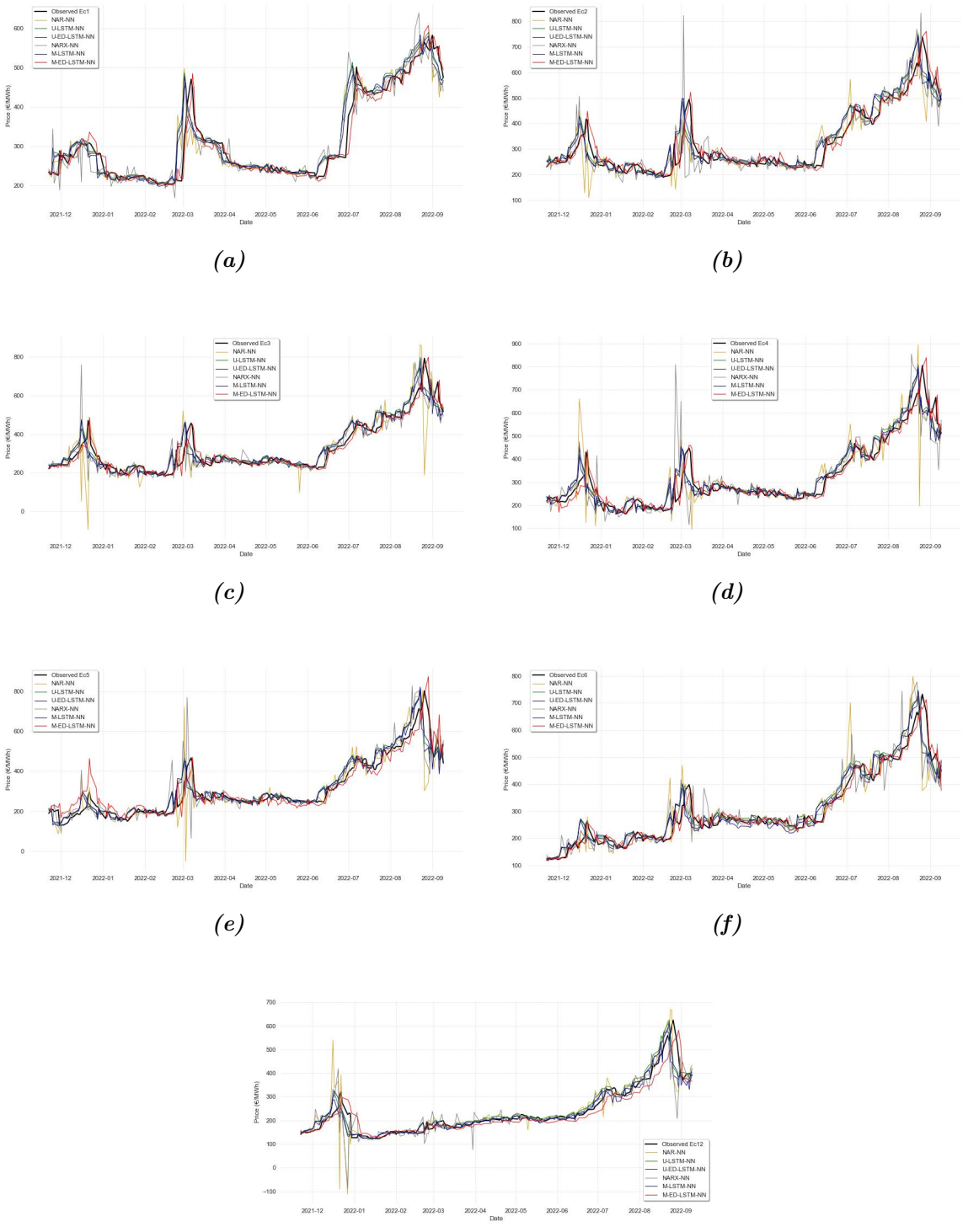
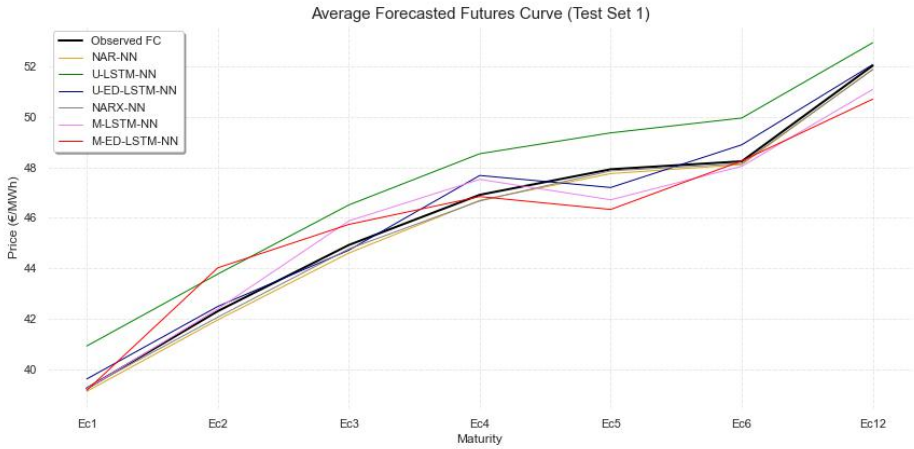


Figure 4.5 *– Models point forecasts of the price series of the Ec1 (a), Ec2 (b), Ec3 (c), Ec4 (d), Ec5 (e), Ec6 (f) and Ec12 (g) futures contracts within Test Set 1.*

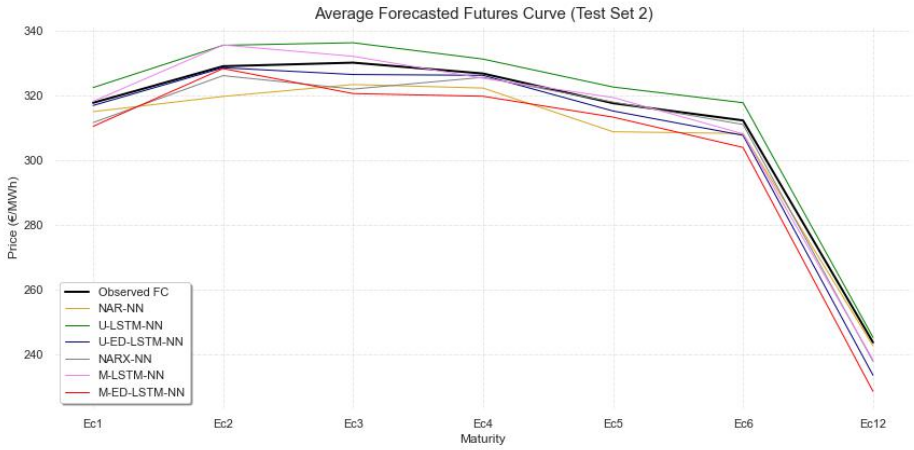


(a) ***(b)*** ***(c)*** ***(d)*** ***(e)*** ***(f)*** ***(g)***

Figure 4.6 – *Models point forecasts of the price series of the Ec1 (a), Ec2 (b), Ec3 (c), Ec4 (d), Ec5 (e), Ec6 (f) and Ec12 (g) futures contracts within Test Set 2.*



(a)



(b)

Figure 4.7 *– Comparison of the Test Set 1 (a) and Test Set 2 (b) average observed futures curves with the average forecasted ones with different neural network models.*

4.5 Conclusion

In this paper, we provided a univariate and multivariate deep learning framework for predicting the term structure of electricity futures prices using Recurrent Neural Networks (RNNs). Specifically, our approach utilized the NAR–NN, NARX–NN, univariate/multivariate LSTM–NN and univariate/multivariate ED–LSTM–NN models to capture the nonlinear temporal relationship of the data as well as complex hidden correlation features with various exogenous factors. The feasibility of the approach

was investigated using two different time spans, that is a stable period and a period of global instability characterized by higher volatility and price peaks which were used as test sets to analyse and compare the models point forecasts accuracy.

The verification experiment showed that the proposed methods can effectively manage trends and dynamics characterizing electricity futures curves and thus provide their effective day–ahead predictions with high levels of accuracy. The comparative analysis indicate that the NAR–NN model performed best only within the more stable period. The LSTM–based models, on the other hand side, exhibited good performances in Test Set 1 and, more important, superior and stable results in Test Set 2. Moreover, the use of two distinct test sets allowed to validate the overall reliability and stability of the proposed framework. Despite the good results which have been obtained up to now, there exists room for improvements. Futures research may include the testing of alternative model configurations and combinations as well as examining the impact of different exogenous variables. Actually all these topics represent a part of our ongoing research.

4.6 Appendix

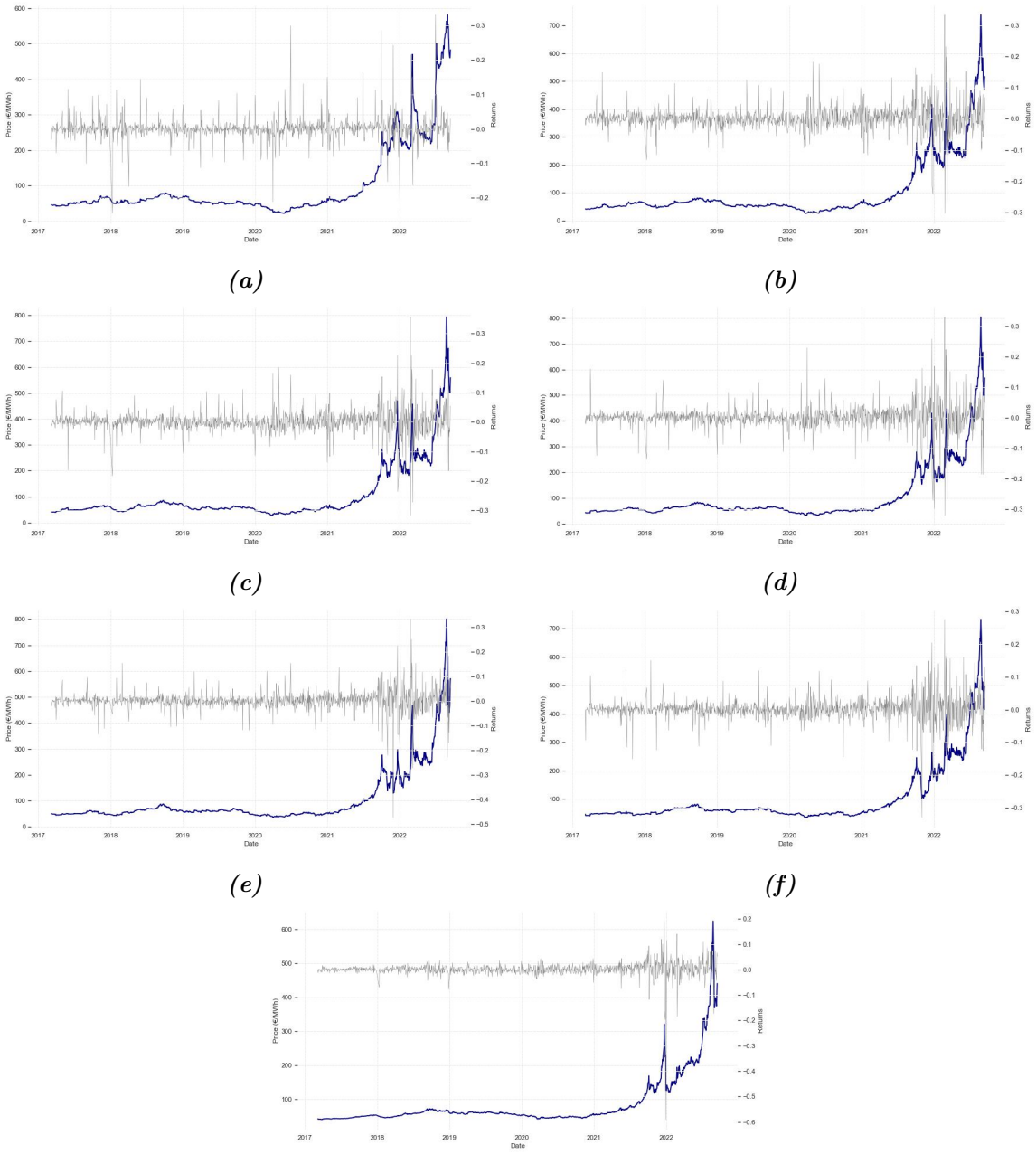


Figure 4.8 – *Time series of daily closing prices (blue) and returns (grey) on the Ec1 (a), Ec2 (b), Ec3 (c), Ec4 (d), Ec5 (e), Ec6 (f) and Ec12 (g) futures contracts.*

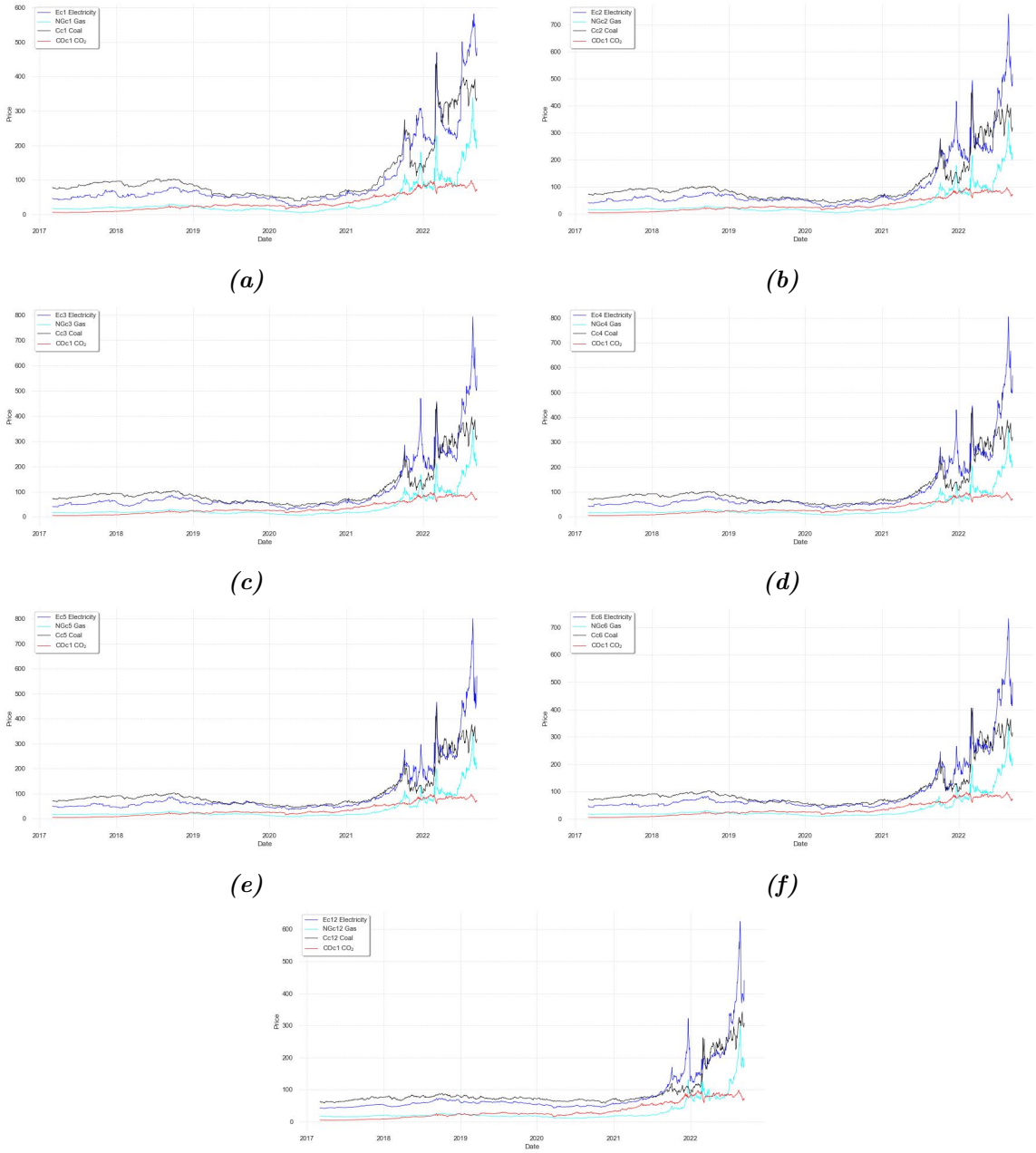


Figure 4.9 – *Time series of Electricity (dark blue), Natural Gas (light blue), Coal (black) futures prices and Carbon Certificate Emissions (CO_2) spot prices (red). The related maturities are indicated in each plots legend.*

Conclusion

This dissertation examined parametric and non-parametric techniques to model and forecast the yield curve in the BRICS bond market as well as futures curves in the Natural Gas and Electricity derivatives markets. The contribution of the thesis is threefold. First, we introduced a technique to determine optimal time–varying decay factors and parameters in parametric models very suitable to manage the yield curve of emerging economies, with a focus on BRICS. Our aim was firstly to endow the existing models with greater flexibility to manage the challenging dynamics observed on these markets. Furthermore, we tested the ability of the improved methods to deliver reliable predictions, under various market conditions, when combined to AR(1), TBATS and ARIMA–NARNN used as local data generating process. We showed that the proposed framework allowed not only a significant improvement of models fitting power, but also make stable and accurate yield curve forecasts.

In a second moment we turned the attention on the Natural Gas futures market and we provided evidence of similarity patterns with fixed–income markets in terms of curves structure and statistical features. The findings allowed to formulate a novel ensemble approach for fitting and forecasting of futures curves trends by using the Four-Factor Dynamic Nelson-Siegel-Svensson (4F-DNSS), the Five-Factor Dynamic De Rezende-Ferreira (5F-DRF), the B-Spline and the NAR-NN models. All models demonstrated high degree of flexibility and adaptability to a wide variety of situations characterizing the market. Nevertheless, the B–Spline achieved undoubtedly highest results properly replicating all the observed curve shapes. Furthermore, we highlighted that the joint use of B–Splines and NAR–NN is the preferable approach for day–ahead forecasting.

We ended the dissertation with a focus on electricity futures curves. We examined a battery of Dynamic Recurrent Neural Networks: NAR–NN, NARX–NN, LSTM–NN and ED–LSTM–NN using both univariate and multivariate settings with and without exogenous variables aimed to identify the most appropriate forecasting strategy. The

predictive power of the models was investigated using two different test sets which demonstrated the reliable accuracy and the robustness of the methodology.

Overall we can draw some conclusions:

1. Existing parametric models can benefit of dynamic adjustment of parameters that make them more reliable both in sample and out of sample when applied to turbulent markets.

2. There are similarity patterns between the fixed–income and the Natural Gas futures markets that make it straightforward to apply on the latter improved parametrics models. Additionally, we highlight that combining parametric models to machine learning generates considerable better results in forecasting.

3. The spiky nature of electricity prices is well integrated by the Dynamic Recurrent Neural Networks.

Bibliography

Aliberti, A., Pupillo, I., Terna, S., Macii, E., Di Cataldo, S., Patti, E., and Acquaviva, A. (2019). A Multi-Patient Data-Driven Approach to Blood Glucose Prediction. *IEEE Access*, 7:69311–69325.

Almansour, A. (2016). Convenience yield in commodity price modeling: A regime switching approach. *Energy Economics*, 53:238–247. Energy Markets.

Apergis, N., Gozgor, G., Lau, C., and Wang, S. (2019). Decoding the Australian electricity market: New evidence from three-regime hidden semi-Markov model. *Energy Economics*, 78:129–142.

Barth, A. and Benth, F. (2014). The forward dynamics in energy markets – infinite-dimensional modelling and simulation. *Stochastics*, 86(6):932–966.

Bartuska, V., Lang, and Nosko (2019). The Geopolitics of Energy Security in Europe. Carnegie reports, Carnegie Europe.

Baruník, J. and Malinská, B. (2016). Forecasting the term structure of crude oil futures prices with neural networks. *Applied Energy*, 164:366–379.

Behmiri, N., Manera, M., and Nicolini, M. (2019). Understanding Dynamic Conditional Correlations between Oil, Natural Gas and Non-Energy Commodity Futures Markets. *The Energy Journal*, 40(2):55–76.

Bekiros, S. and Avdoulas, C. (2020). Revisiting the Dynamic Linkages of Treasury Bond Yields for the BRICS: A Forecasting Analysis. *Forecasting*, 2(2):102–129.

Benrhmach, G., Namir, K., Namir, A., and Bouyaghroumni, J. (2020). Nonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time Series. *Journal of Applied Mathematics*, 2020.

Benth, F. and Koekebakker, S. (2008). Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157.

Berrisch, J. and Ziel, F. (2020). Modeling and Probabilistic Forecasting of Natural Gas Prices.

Bessembinder, H., Coughenour, J., Seguin, P., and Smoller, M. (1996). Is There a Term Structure of Futures Volatilities? Reevaluating the Samuelson Hypothesis. *The Journal of Derivatives*, 4(2):45–58.

Bhattarai, S., Chatterjee, A., and Park, W. (2021). Effects of US quantitative easing on emerging market economies. *Journal of Economic Dynamics and Control*, 122:104031.

Biagini, F., Bregman, J., and Meyer-Brandis, T. (2015). Electricity Futures Price Modeling with Lévy Term Structure Models. *International Journal of Theoretical and Applied Finance*, 18(01):1550003.

Blanchard, T. and Samanta, B. (2020). Wind speed forecasting using neural networks. *Wind Engineering*, 44(1):33–48.

Bliss, R. (1996). Testing term structure estimation methods. Working Paper 96-12a, Federal Reserve Bank of Atlanta.

Bordoff, J. and O'Sullivan, M. (2023). The Age of Energy Insecurity. How the Fight for Resources Is Upending Geopolitics. *Foreign Affairs*, 102(2).

Borovkova, S. and Mahakena, D. (2015). News, volatility and jumps: the case of natural gas futures. *Quantitative Finance*, 15(7):1217–1242.

Borovkova, S. and Schmeck, M. (2017). Electricity price modeling with stochastic time change. *Energy Economics*, 63:51–65.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2016). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics, fifth edition.

Brigida, M. (2014). The switching relationship between natural gas and crude oil prices. *Energy Economics*, 43:48–55.

Brown, S. and Yücel, M. (2008). What Drives Natural Gas Prices? *The Energy Journal*, 29(2):45–60.

Brownlee, J. (2018). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

Butler, S., Kokoszka, P., Miao, H., and Shang, H. (2021). Neural network prediction of crude oil futures using B-splines. *Energy Economics*, 94:105080.

Caldeira, J., Gupta, R., Suleman, M., and Torrent, H. (2020). Forecasting the Term Structure of Interest Rates of the BRICS: Evidence from a Nonparametric Functional Data Analysis. *Emerging Markets Finance and Trade*, 0(0):1–18.

Caldeira, J., Moura, G., and Portugal, M. (2010). Efficient Yield Curve Estimation and Forecasting in Brazil. *Revista EconomiA*, 11(1).

Caldeira, J., Moura, G., Santos, A., and Tourrucôo, F. (2016). Forecasting the yield curve with the arbitrage-free dynamic Nelson–Siegel model: Brazilian evidence. *EconomiA*, 17(2):221–237.

Castello, O. and Resta, M. (2019). *DeRezende.Ferreira: Zero Coupon Yield Curve Modelling*. University of Genova - Department of Economics and Business Studies. R package version 0.1.0.

Castello, O. and Resta, M. (2022). Modeling the Yield Curve of BRICS Countries: Parametric vs. Machine Learning Techniques. *Risks*, 10(2).

Chadha, J., Durre, A., Sarno, L., and Joyce, M. (2014). *Developments in Macro-Finance Yield Curve Modelling*. Cambridge University Press.

Chakroun, F. and Abid, F. (2014). A Methodology to Estimate the Interest Rate Yield Curve in Illiquid Market: The Tunisian Case. *Journal of Emerging Market Finance*, 13(3):305–333.

Chen, B., Xiong, R., Li, H., Sun, Q., and Yang, J. (2019). Pathways for sustainable energy transition. *Journal of Cleaner Production*, 228:1564–1571.

Chen, Y., Chua, W., and Koch, T. (2018). Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. *Applied Energy*, 228:1091–1110.

Cheng, H., Ding, X., Zhou, W., and Ding, R. (2019). A hybrid electricity price forecasting model with Bayesian optimization for German energy exchange. *International Journal of Electrical Power & Energy Systems*, 110:653–666.

Cheng, I. and Xiong, W. (2014). The Financialization of Commodity Markets. Working Paper 19642, National Bureau of Economic Research.

Chi, Y. (2021). Time Series Forecasting of Global Price of Soybeans using a Hybrid SARIMA and NARNN Model: Time Series Forecasting of Global Price of Soybeans. *Data Science: Journal of Computing and Applied Informatics*, 5(2):85–101.

Chiarella, C., Clewlow, L., and Kang, B. (2009). Modelling and estimating the forward price curve in the energy market. Technical Report 260, University of Technology Sydney.

Chiţu, L., Quint, D., et al. (2018). Emerging market vulnerabilities - a comparison with previous crises. Technical report, European Central Bank.

Chollet, F. et al. (2015). Keras. https://keras.io.

Chou, J., Su, Y., Tang, H., and Chen, C. (2009). Fitting the term structure of interest rates in illiquid market: Taiwan experience. *Investment Management and Financial Innovations*, 6, Issue 1:101–116.

Chouikh, A., Yousfi, R., and Chehibi, C. (2017). Yield Curve Estimation: An Empirical Evidence from the Tunisian Bond Market. *Journal of Finance and Economics*, 5(6):300–309.

Christensen, J., Diebold, F., and Rudebusch, G. (2007). The Affine Arbitrage–Free Class of Nelson–Siegel Term Structure Models. Working Paper 13611, National Bureau of Economic Research.

Christensen, J., Diebold, F., and Rudebusch, G. (2009). An Arbitrage-Free Generalized Nelson–Siegel Term Structure Model. *The Econometrics Journal*, 12(3):C33–C64.

Conejo, A., Plazas, M., Espinola, R., and Molina, A. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, 20(2):1035–1042.

Copernicus Climate Change Service (C3S) (2022). Summer 2022 Europe's hottest on record.

Coulon, M. and Howison, S. (2009). Stochastic behavior of the electricity bid stack: from fundamental drivers to power prices. *The Journal of Energy Markets*, 2(1):29–69.

Creti, A. and Nguyen, D. (2015). Energy markets financialization, risk spillovers, and pricing models. *Energy Policy*, 82:260–263.

Curry, H. and Schoenberg, I. (1947). On spline distributions and their limits: The Polya Distribution Functions. *Bulletin of the American Mathematical Society*, 4:109.

Curry, H. and Schoenberg, I. (1966). On Pólya frequency functions IV: The fundamental spline functions and their limits. *Journal d'Analyse Mathématique*, 17:71 – 107.

da Silva, P. and Horta, P. (2019). The effect of variable renewable energy sources on electricity price volatility: the case of the Iberian market. *International Journal of Sustainable Energy*, 38(8):794–813.

de Boyrie, M. and Pavlova, I. (2016). Dynamic interdependence of sovereign credit default swaps in BRICS and MIST countries. *Applied Economics*, 48(7):563–575.

de Marcos, R., Bunn, D., Bello, A., and Reneses, J. (2020). Short-Term Electricity Price Forecasting with Recurrent Regimes and Structural Breaks. *Energies*, 13(20).

De Pooter, M. (2007). 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.

De Pooter, M., Ravazzolo, F., and van Dijk, D. (2010). Term structure forecasting using macro factors and forecast combination. International Finance Discussion Papers 993, Board of Governors of the Federal Reserve System (U.S.).

De Rezende, R. and Ferreira, M. (2008). 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.

De Rezende, R. and Ferreira, M. (2013). Modeling and Forecasting the Yield Curve by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach. *Journal of Forecasting*, 32(2):111 – 123.

Deng, J., Song, W., and Zio, E. (2020). A Discrete Increment Model for Electricity Price Forecasting Based on Fractional Brownian Motion. *IEEE Access*, 8:130762–130770.

Deschatre, T., Féron, O., and Gruet, P. (2021). A survey of electricity spot and futures price models for risk management applications. *Energy Economics*, 102:105504.

Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, 7(3):1174–1179.

Di Franco, G. and Santurro, M. (2020). Machine learning, artificial neural networks and social research. *Quality & Quantity*, 55(3):1007 – 1025.

Diebold, F. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X., Rudebusch, G. D., and Borağan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1):309–338.

Diebold, F. X. and Rudenbusch, G. (2017). *Yield Curve Modeling and Forecasts*. Princeton University Press.

Dillig, M., Jung, M., and Karl, J. (2016). The impact of renewables on electricity prices in Germany – An estimation based on historic spot prices in the years 2011–2013. *Renewable and Sustainable Energy Reviews*, 57:7–15.

Dolatabadi, S., Narayan, P., Nielsen, M., and Xu, K. (2018). Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. *Journal of Futures Markets*, 38(2):219–242.

Dong, Y., Wang, J., Jiang, H., and Wu, J. (2011). Short-term electricity price forecast based on the improved hybrid model. *Energy Conversion and Management*, 52(8):2987–2995.

Dudek, G. (2016). Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32(3):1057–1060.

Duong, H. and Kalev, P. (2008). The Samuelson hypothesis in futures markets: An analysis using intraday data. *Journal of Banking & Finance*, 32(4):489–500.

El-Shagi, M. and Jiang, L. (2019). Efficient Dynamic Yield Curve Estimation in Emerging Financial Markets. CFDS Discussion Paper Series 2019/4, Center for Financial Development and Stability at Henan University, Kaifeng, Henan, China.

Emery, G. and Liu, Q. (2002). An analysis of the relationship between electricity and natural-gas futures prices. *Journal of Futures Markets*, 22(2):95–122.

Ertan, A., Karahan, C., and Temuçin, T. (2020). Estimating the Yield Curve for Sovereign Bonds: The Case of Turkey. *Finans Politik & Ekonomik Yorumlar*, 57(653):137–159.

Etienne, X., Trujillo-Barrera, A., and Wiggins, S. (2016). Price and volatility transmissions between natural gas, fertilizer, and corn markets. *Agricultural Finance Review*, 76(1):151–171.

Eydeland, A. and Wolyniec, K. (2002). *Energy and power risk management: New developments in modeling, pricing, and hedging*, volume 97. John Wiley & Sons.

Fanelli, V., Maddalena, L., and Musti, S. (2016). Modelling electricity futures prices using seasonal path-dependent volatility. *Applied Energy*, 173:92–102.

Faria, A. and Almeida, C. (2018). A hybrid spline-based parametric model for the yield curve. *Journal of Economic Dynamics and Control*, 86:72–94.

Fałdziński, M., Fiszeder, P., and Orzeszko, W. (2021). Forecasting Volatility of Energy Commodities: Comparison of GARCH Models with Support Vector Regression. *Energies*, 14(1).

Fernandes, M. and Vieira, F. (2019). A dynamic Nelson–Siegel model with forward-looking macroeconomic factors for the yield curve in the US. *Journal of Economic Dynamics and Control*, 106:103720.

Filipović, D. (2009). *Term Structure Models*. Springer-Verlag.

Furtuna, O., Grassi, A., Ianiro, A., Kallage, K., Koci, R., Lenoci, F. D., Sowinski, A., and Vacirca, F. (2022). Financial stability risks from energy derivatives markets. techreport, European Central Bank.

Füss, R., Mahringer, S., and Prokopczuk, M. (2015). Electricity derivatives pricing with forward-looking information. *Journal of Economic Dynamics and Control*, 58:34–57.

Futures Industry Association (FIA) (2022). 2021 Annual ETD Volume Review.

Gas Infrastructure Europe (GIE) (2022). Gas Storage Inventory.

Gatfaoui, H. (2016). Linking the gas and oil markets with the stock market: Investigating the U.S. relationship. *Energy Economics*, 53:5–16. Energy Markets.

Ghosh, S., Bohra, A., and Dutta, S. (2021). The Texas Freeze of February 2021: Event and Winterization Analysis Using Cost and Pricing Data. In *2021 IEEE Electrical Power and Energy Conference (EPEC)*, pages 7–13.

Gibson, R. and Schwartz, E. (1990). Stochastic Convenience Yield and the Pricing of Oil Contingent Claims. *The Journal of Finance*, 45(3):959–976.

Gonzalez, V., Contreras, J., and Bunn, D. (2012a). Forecasting Power Prices Using a Hybrid Fundamental-Econometric Model. *IEEE Transactions on Power Systems*, 27(1):363–372.

Gonzalez, V., Contreras, J., and Bunn, D. (2012b). Forecasting Power Prices Using a Hybrid Fundamental-Econometric Model. *IEEE Transactions on Power Systems*, 27(1):363–372.

Gudkov, N. and Ignatieva, K. (2021). Electricity price modelling with stochastic volatility and jumps: An empirical investigation. *Energy Economics*, 98:105260.

Guidolin, M. and Alpcan, T. (2019). Transition to sustainable energy generation in Australia: Interplay between coal, gas and renewables. *Renewable Energy*, 139:359–367.

Guo, K., Liu, F., Sun, X., Zhang, D., and Ji, Q. (2023). Predicting natural gas futures' volatility using climate risks. *Finance Research Letters*, page 103915.

Guo, W. and Xue, H. (2014). Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. *Mathematical Problems in Engineering*, 2014.

Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the Term Structure. *Journal of Economic Literature*, 50(2):331–67.

Heath, D., Jarrow, R., and Morton, A. (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77–105.

Heather, P. (2021). European Traded Gas Hubs: German hubs about to merge. techreport, The Oxford Institute for Energy Studies.

Hess, M. (2020). A pure-jump mean-reverting short rate model. *Modern Stochastics: Theory and Applications*, 7(2):113–134.

Higgs, H. and Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. *Energy Economics*, 30(6):3172–3185. Technological Change and the Environment.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hoffmaister, A., Roldós, J., and Tuladhar, A. (2010). Yield curve dynamics and spillovers in central and eastern European countries. resreport 051, International Monetary Fund.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Horváth, L., Liu, Z., Rice, G., and Wang, S. (2020). A functional time series analysis of forward curves derived from commodity futures. *International Journal of Forecasting*, 36(2):646–665.

Hribar, R., Potočnik, P., Šilc, J., and Papa, G. (2019). A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy*, 167:511–522.

Hyndman, R. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3):1 – 22.

Intercontinental Exchange (2021). ICE Announces Record Activity in TTF and JKM Gas Complexes as They Evolve into Global Benchmarks. techreport, Intercontinental Exchange (ICE).

Islyaev, S. and Date, P. (2015). Electricity futures price models: Calibration and forecasting. *European Journal of Operational Research*, 247(1):144–154.

Itoh, Y. and Kobayashi, T. (2010). A single-factor model analysis of the electricity futures price and its application. *Electrical Engineering in Japan*, 172(4):1–9.

Jablonowski, C. and Schicks, M. (2017). A Three-Factor Model on the Natural Gas Forward Curve Including Temperature Forecasts. *Journal of Energy Markets*, 10(3).

Jaeck, E. and Lautier, D. (2014). Samuelson hypothesis and electricity derivative markets. In *31st International French Finance Association Conference, AFFI 2014*, page 24, Aix-en-Provence, France.

Jahangir, H., Tayarani, H., Baghali, S., Ahmadian, A., Elkamel, A., Golkar, M., and Castilla, M. (2020). A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks. *IEEE Transactions on Industrial Informatics*, 16(4):2369–2381.

Jana, R. and Ghosh, I. (2022). A residual driven ensemble machine learning approach for forecasting natural gas prices: analyses for pre-and during-COVID-19 phases. *Annals of Operations Research*.

Ji, Q., Zhang, H., and Geng, J. (2018). What drives natural gas prices in the United States? – A directed acyclic graph approach. *Energy Economics*, 69:79–88.

Jin, J. and Kim, J. (2015). Forecasting Natural Gas Prices Using Wavelets, Time Series, and Artificial Neural Networks. *PLOS ONE*, 10(11):1–23.

Jonckheere, A. (1954). A Distribution-Free k-Sample Test Against Ordered Alternatives. *Biometrika*, 41(1/2):133–145.

Kallabis, T., Pape, C., and Weber, C. (2016). The plunge in German electricity futures prices – Analysis using a parsimonious fundamental model. *Energy Policy*, 95:280–290.

Kanamura, T. and Bunn, D. (2022). Market making and electricity price formation in Japan. *Energy Economics*, 107:105765.

Kang, K. (2012). Forecasting the Term Structure of Korean Government Bond Yields Using the Dynamic Nelson-Siegel Class Models. *Asia-Pacific Journal of Financial Studies*, 41(6):765–787.

Karakatsani, N. and Bunn, D. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4):764–785. Energy Forecasting.

Karstanje, D., Van Der Wel, M., and van Dijk, D. (2017). Common Factors in Commodity Futures Curves. *Available at SSRN 2558014*.

Kavitha, G. and Kalpana, K. (2023). Integrated Tuning of Hidden Markov Parametric Optimization Model with Genetic Algorithm for Electricity Market Forecasting. In Silhavy, R., Silhavy, P., and Prokopova, Z., editors, *Data Science and Algorithms in Systems*, pages 798–806, Cham. Springer International Publishing.

Khan, M. A. (2015). Modelling and forecasting the demand for natural gas in Pakistan. *Renewable and Sustainable Energy Reviews*, 49:1145–1159.

Khudoley, K. (2022). New Russia-West Confrontation: War of Attrition or Escalation? *Strategic Analysis*, 46(6):571–584.

Kim, W. J., Jung, G., and Choi, S. Y. (2020). Forecasting CDS Term Structure Based on Nelson-Siegel Model and Machine Learning. *Complexity*, 2020:2518283:1–2518283:23.

Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Koekebakker, S. and Ollmar, F. (2005). Forward curve dynamics in the Nordic electricity market. *Managerial Finance*, 31(6):73–94.

Koopman, S. J., Mallee, M. I. P., and van der Wel, M. (2007). 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.

Kostrzewski, M. and Kostrzewska, J. (2019). Probabilistic electricity price forecasting with Bayesian stochastic volatility models. *Energy Economics*, 80:610–620.

Kumar, V., Singh, N., Singh, D., and Mohanty, S. (2018). Short-Term Electricity Price Forecasting Using Hybrid SARIMA and GJR-GARCH Model. In Perez, G. M., Mishra, K. K., Tiwari, S., and Trivedi, M. C., editors, *Networking Communication and Data Knowledge Engineering*, pages 299–310, Singapore. Springer Singapore.

Kuo, P. and Huang, C. (2018). An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. *Sustainability*, 10(4).

Kwas, M. and Rubaszek, M. (2021). Forecasting Commodity Prices: Looking for a Benchmark. *Forecasting*, 3(2):447–459.

Lago, J., Marcjasz, G., De Schutter, B., and Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983.

Lantz, B. (2019). *Machine Learning with R: Expert techniques for predictive modeling*. Packt Publishing, 3 edition.

Lartey, V. and Li, Y. (2018). Zero-Coupon and Forward Yield Curves for Government of Ghana Bonds. *SAGE Open*, 8(3):2158244018800786.

Lartey, V. C., Li, Y., Lartey, H. D., and Boadi, E. K. (2019). Zero-Coupon, Forward, and Par Yield Curves for the Nigerian Bond Market. *SAGE Open*, 9(4):2158244019885144.

Leonhardt, D., Ware, A., and Zagst, R. (2017). A Cointegrated Regime-Switching Model Approach with Jumps Applied to Natural Gas Futures Prices. *Risks*, 5(3).

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

Li, B. (2019). Pricing dynamics of natural gas futures. *Energy Economics*, 78:91–108.

Li, H., Chen, L., Wang, D., and Zhang, H. (2017). Analysis of the Price Correlation between the International Natural Gas and Coal. *Energy Procedia*, 142:3141–3146. Proceedings of the 9th International Conference on Applied Energy.

Li, J., Wu, Q., Tian, Y., and Fan, L. (2021). Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network. *Energy*, 227:120478.

Li, R. and Song, X. (2023). A multi-scale model with feature recognition for the use of energy futures price forecasting. *Expert Systems with Applications*, 211:118622.

Liang, C., Xia, Z., Lai, X., and Wang, L. (2022). Natural gas volatility prediction: Fresh evidence from extreme weather and extended GARCH-MIDAS-ES model. *Energy Economics*, 116:106437.

Lin, B. (2002). Fitting term structure of interest rates using B-splines: the case of Taiwanese Government bonds. *Applied Financial Economics*, 12(1):57–75.

Linton, O., Mammen, E., Nielsen, J., and Tanggaard, C. (2001). Yield curve estimation by kernel smoothing methods. *Journal of Econometrics*, 105(1):185–223. Forecasting and empirical methods in finance and macroeconomics.

Lopez De Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley, New York.

Lorenčič, E. (2016). Testing the Performance of Cubic Splines and Nelson-Siegel Model for Estimating the Zero-coupon Yield Curve. *Naše gospodarstvo/Our economy*, 62(2):42–50.

Lu, F., Ma, F., Li, P., and Huang, D. (2022). Natural gas volatility predictability in a data-rich world. *International Review of Financial Analysis*, 83:102218.

Luo, D., Pang, T., and Xu, J. (2021). Forecasting U.S. Yield Curve Using the Dynamic Nelson–Siegel Model with Random Level Shift Parameters. *Economic Modelling*, 94:340–350.

Lv, X. and Shan, X. (2013). Modeling natural gas market volatility using GARCH with different distributions. *Physica A: Statistical Mechanics and its Applications*, 392(22):5685–5699.

Maciejowska, K., Nitka, W., and Weron, T. (2021). Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Economics*, 99:105273.

Marquardt, D. (1963a). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Marquardt, D. W. (1963b). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Mason, C. and N.A. Wilmot (2014). Jump processes in natural gas markets. *Energy Economics*, 46:S69–S79. Special Issue on Recent Approaches to Modelling Oil and Energy Commodity Prices.

Mehrdoust, F. and Noorani, I. (2021). Forward price and fitting of electricity Nord Pool market under regime-switching two-factor model. *Mathematics and Financial Economics*, 15(3):501–543.

Mineo, E., Alencar, A., Moura, M., and Fabris, A. (2020). Forecasting the Term Structure of Interest Rates with Dynamic Constrained Smoothing B-Splines. *Journal of Risk and Financial Management*, 13(4):119 – 132.

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, 29(1):46–63.

Muthoni, L., Onyango, S., and Ongati, O. (2015). Extraction of Zero-Coupon Yield Curve for Nairobi Securities Exchange: Finding the Best Parametric Model for East African Securities Markets. *Journal of Mathematics and Statistical Science*, 2015:51–74.

Muvingi, J. and Kwinjo, T. (2014). Estimation of term Structures using Nelson-Siegel and Nelson-Siegel-Svensson: A Case of a Zimbabwean Bank. *Journal of Applied Finance & Banking*, 4(6):155–190.

Nagy, K. (2020). Term structure estimation with missing data: Application for emerging markets. *The Quarterly Review of Economics and Finance*, 75:347–360.

Najafi, A., Taleghani, R., and Mehrdoust, F. (2023). Forward contract prices of electricity Nord Pool market: calibration and jump approximation. *Sādhanā*, 48(1):11–48.

Nascimento, J., Pinto, T., and Vale, Z. (2019). Electricity Price Forecast for Futures Contracts with Artificial Neural Network and Spearman Data Correlation. In Rodríguez, S., Prieto, J., Faria, P., Kłos, S., Fernández, A., Mazuelas, S., Jiménez-López, M. D., Moreno, M. N., and Navarro, E. M., editors, *Distributed Computing and Artificial Intelligence, Special Sessions, 15th International Conference*, pages 12–20, Cham. Springer International Publishing.

Nelson, C. and Siegel, A. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473–89.

Nowotarski, J. and Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568.

Nunes, M., Gerding, E., McGroarty, F., and Niranjan, M. (2019). A comparison of multitask and single task learning with artificial neural networks for yield curve forecasting. *Expert Systems with Applications*, 119:362–375.

Oduor, D. (2022). Mean: Reverting logistic Brownian motion with jump diffusion process on energy commodity prices. *International Journal of Statistics and Applied Mathematics*, 7(4):69–74.

Olivares, K., Challu, C., Marcjasz, G., Weron, R., and Dubrawski, A. (2022). Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *International Journal of Forecasting*.

Ö zmen, A., Yılmaz, Y., and Weber, G. (2018). Natural gas consumption forecast with MARS and CMARS models for residential users. *Energy Economics*, 70:357–381.

Panapakidis, I. and Dagoumas, A. (2017). Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy*, 118:231–245.

Panella, M., Barcellona, F., and D'Ecclesia, R. (2012). Forecasting Energy Commodity Prices Using Neural Networks. *Advances in Decision Sciences*, 2012.

Pascual, C. and Zambetakis, E. (2016). The geopolitics of energy: From security to survival. Technical Report 07, Brookins Institution.

Pei, Y., Huang, C., Shen, Y., and Wang, M. (2023). A Novel Model for Spot Price Forecast of Natural Gas Based on Temporal Convolutional Network. *Energies*, 16(5).

Peng, L., Liu, S., Liu, R., and Wang, L. (2018). Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy*, 162:1301–1314.

Pereda, J. (2009). Estimacion de la Curva de Rendimiento Cupon Cero para el Perú. Technical Report 17, Banco Central De Reserva Del Perú.

Petousis, T. and Barr, G. (2016). Nelson Siegel Parameterisation of the South African Sovereign Yield Curve: A Link to the Rand Exchange Rate and Jse Sectors. *Studies in Economics and Econometrics*, 40(1):41–70.

Pircalabu, A. and Benth, F. (2017). A regime-switching copula approach to modeling day-ahead prices in coupled electricity markets. *Energy Economics*, 68:283–302.

Poghosyan, K. and Poghosyan, A. (2019). Yield Curve Estimation and Forecasting in Armenia. *Armenian Journal of Economics*, 4:1–21.

Posthaus, A. (2019). Yield Curve Fitting with Artificial Intelligence: A Comparison of Standard Fitting Methods with AI Algorithms. *Journal of Computational Finance*, 22(5):1 – 23.

Prasanna, K. and Sowmya, S. (2017). Yield curve in India and its interactions with the US bond market. *International Economics and Economic Policy*, 14(2):353–375.

Rafiei, M., Niknam, T., and Khooban, M. (2017). Probabilistic electricity price forecasting by improved clonal selection algorithm and wavelet preprocessing. *Neural Computing and Applications*, 28(12):3889–3901.

Raviv, E., Bouwman, K. E., and van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, 50:227–239.

Reed, R. D. and Marks, R. J. (1999). *Neural Smithing: Supervise Learning in Feedforward Artificial Neural Networks*. The MIT Press.

Ribeiro, M., Stefenon, S., de Lima, J., Nied, A., Mariani, V., and Coelho, L. (2020). Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. *Energies*, 13(19).

Rosadi, D., Nugraha, Y., and Dewi, R. (2011). Forecasting the Indonesian Government Securities Yield Curve using Neural Networks and Vector Autoregressive Model. Technical report, Department of Mathematics, Gadjah Mada University, Indonesia.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological review*, 65 6:386–408.

R.R. Wahlstrom, F. Paraschiv, and M. Schurle (2021). A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. *Computational Economics*.

Ruiz, L., Cuéllar, M., Calvo-Flores, M., and Jiménez, M. (2016). An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies*, 9(9).

Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.

Salehnia, N., Falahi, M., Seifi, A., and Mahdavi Adeli, M. (2013). Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *Journal of Natural Gas Science and Engineering*, 14:238–249.

Salisu, A., Cuñado, J., Isah, K., and Gupta, R. (2021). Stock markets and exchange rate behaviour of the BRICS. *Journal of Forecasting*, 40(8):1581 – 1595.

Saltik, O., Degirmen, S., and Ural, M. (2016). Volatility Modelling in Crude Oil and Natural Gas Prices. *Procedia Economics and Finance*, 38:476–491. The 5th Istanbul Conference of Economics and Finance.

Samuelson, P. A. (1973). Proof That Properly Discounted Present Values of Assets Vibrate Randomly. *The Bell Journal of Economics and Management Science*, 4(2):369–374.

Saunders, A. and Cornett, M. (2014). *Financial Markets and Institutions, Sixth Edition*. The McGraw-Hill/Irwin Series in Finance, Insuarance, and Real Estate.

Schoenberg, I. (1946). Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions: Paert A. - On the Problem of Smoothing or Graduation. A First Class of Analytic Approximation Fomulae. *Quarterly of Applied Mathematics*, 4(1):45–99.

Shaikh, F. and Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, 77:25–32.

Shaikh, F., Ji, Q., Shaikh, P. H., Mirjat, N. H., and Uqaili, M. A. (2017). Forecasting China's natural gas demand based on optimised nonlinear grey models. *Energy*, 140:941–951.

Snam (2020). Global Gas Report 2020.

Stuart, R. (2020). The term structure, leading indicators, and recessions: evidence from Switzerland, 1974–2017. *Swiss Journal of Economics and Statistics*, 156(2).

Su, H., Peng, X., Liu, H., Quan, H., Wu, K., and Chen, Z. (2022). Multi-Step-Ahead Electricity Price Forecasting Based on Temporal Graph Convolutional Network. *Mathematics*, 10(14).

Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. *Energy*, 178:585–597.

Suenaga, H., Smith, A., and Williams, J. (2008). Volatility dynamics of NYMEX natural gas futures prices. *Journal of Futures Markets*, 28(5):438–463.

Suimon, Y., Sakaji, H., Izumi, K., and Matsushima, H. (2020). Autoencoder-Based Three-Factor Model for the Yield Curve of Japanese Government Bonds and a Trading Strategy. *Journal of Risk and Financial Management*, 13(4):1 – 21.

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to Sequence Learning with Neural Networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. NBER Working Papers 4871, National Bureau of Economic Research, Inc.

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208–220.

Tan, Z., Zhang, J., Wang, J., and Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, 87(11):3606–3610.

Tang, Y., Wang, Q., Xu, W., Wang, M., and Wang, Z. (2019). Natural Gas Price Prediction with Big Data. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 5326–5330.

Terpstra, T. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae (Proceedings)*, 55:327–333.

Tiwari, A., Mukherjee, Z., Gupta, R., and Balcilar, M. (2019). A wavelet analysis of the relationship between oil and natural gas prices. *Resources Policy*, 60:118–124.

Tschora, L., Pierre, E., Plantevit, M., and Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, 313:118752.

Ullah, W. (2017). Term structure forecasting in affine framework with time-varying volatility. *Stat Methods Appl*, 26:453–483.

Ullah, W. and Bari, K. (2018). The Term Structure of Government Bond Yields in an Emerging Market. *Romanian Journal for Economic Forecasting*, 21(3):5–28.

Umar, Z., Riaz, Y., and Aharon, D. Y. (2022). Network connectedness dynamics of the yield curve of G7 countries. *International Review of Economics & Finance*, 79:275–288.

van Goor, H. and Scholtens, B. (2014). Modeling natural gas price volatility: The case of the UK gas market. *Energy*, 72:126–134.

Čeperić, E., Žiković, S., and Čeperić, V. (2017). Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy*, 140:893–900.

Vela, D. (2013). Forecasting Latin–American yield curves: An artificial neural network approach. techreport 761, Banco de la República (Colombia).

Voronin, S., Partanen, J., and Kauranne, T. (2013). A hybrid electricity price forecasting model for the Nordic electricity spot market. *International Transactions on Electrical Energy Systems*, 24(5):736–760.

Wan, C., Xu, Z., Wang, Y., Dong, Z., and Wong, K. (2014). A Hybrid Approach for Probabilistic Forecasting of Electricity Price. *IEEE Transactions on Smart Grid*, 5(1):463–470.

Wang, J., Cao, J., Yuan, S., and Cheng, M. (2021). 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*, 233:121082.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081.

Wilamowski, B. and Irwin, J. (2011). *Intelligent Systems*. CRC Press, second edition.

Xiong, H. and Mamon, R. (2019). A higher-order Markov chain-modulated model for electricity spot-price dynamics. *Applied Energy*, 233-234:495–515.

Xu, X. and Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture*, 184:106120.

Xu, X. and Zhang, Y. (2022). Soybean and Soybean Oil Price Forecasting through the Nonlinear Autoregressive Neural Network (NARNN) and NARNN with Exogenous Inputs (NARNN–X). *Intelligent Systems with Applications*, 13:200061.

Yang, H. and Schell, K. R. (2021). Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. *Applied Energy*, 299:117242.

Yang, W., Sun, S., Hao, Y., and Wang, S. (2022). A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. *Energy*, 238:121989.

Yang, W., Wang, J., Niu, T., and Du, P. (2020). A novel system for multi-step electricity price forecasting for electricity market management. *Applied Soft Computing*, 88:106029.

Yu, J., Zhou, L., and Xia, A. (2008). Power futures price forecasting based on improved wavelet neural network. In *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 521–526.

Zeb, S. and Rashid, A. (2019). Systemic risk in financial institutions of BRICS: measurement and identification of firm-specific determinants. *Risk Management*, 21:243–264.

Zhang, J., Tan, Z., and Wei, Y. (2020a). An adaptive hybrid model for short term electricity price forecasting. *Applied Energy*, 258:114087.

Zhang, K. and Shi, Q. (2009). Power Futures Price Forecasting Based on RBF Neural Network. In *2009 International Conference on Business Intelligence and Financial Engineering*, pages 50–52.

Zhang, R., Li, G., and Ma, Z. (2020b). A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting. *IEEE Access*, 8:143423–143436.

Zhang, X., Wang, J., and Gao, Y. (2019). A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. *Energy Economics*, 81:899–913.

Zhang, Y., Chevallier, J., and Guesmi, K. (2017). "De-financialization" of commodities? Evidence from stock, crude oil and natural gas markets. *Energy Economics*, 68:228–239.

Zhao, Z., Wang, C., Nokleby, M., and Miller, C. (2017). Improving short-term electricity price forecasting using day-ahead LMP with ARIMA models. In *2017 IEEE Power & Energy Society General Meeting*, pages 1–5.

Zhou, L., Zhao, P., Wu, D., Cheng, C., and Huang, H. (2018). Time series model for forecasting the number of new admission inpatients. *BMC Medical Informatics and Decision Making*, 18(39).

Zhou, S., Zhou, L., Mao, M., Tai, H., and Wan, Y. (2019). An Optimized Heterogeneous Structure LSTM Network for Electricity Price Forecasting. *IEEE Access*, 7:108161–108173.

Ziel, F. and Steinert, R. (2018). Probabilistic mid- and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews*, 94:251–266.

Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420.

Zoricic, D. and Orsag, S. (2013). Parametric yield curve modeling in an illiquid and undeveloped financial market. *UTMS Journal of Economics*, 4(3):243–252.