

Please cite the Published Version

Alshater, Muneer M, Kampouris, Ilias, Marashdeh, Hazem, Atayah, Osama F and Banna, Hasanul (2022) Early warning system to predict energy prices: the role of artificial intelligence and machine learning. *Annals of Operations Research*. ISSN 0254-5330

DOI: <https://doi.org/10.1007/s10479-022-04908-9>

Publisher: Springer (part of Springer Nature)

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/630287/>

Additional Information: This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1007/s10479-022-04908-9>

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Early Warning System to Predict Energy Prices: The Role of Artificial Intelligence and Machine Learning

Muneer M. Alshater

Faculty of Business, Emirates College of Technology, Abu Dhabi, United Arab Emirates.

E-mail: muneermaher@gmail.com

Ilias Kampouris

College of Business, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, United Arab Emirates. E-mail: ilias.kampouris@adu.ac.ae

Hazem Marashdeh

College of Business, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, United Arab Emirates.

E-mail: hazem.marashdeh@adu.ac.ae

Osama F. Atayah

College of Business, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, United Arab Emirates. E-mail: Dr.osamaatayah@gmail.com

Hasanul Banna

Department of Accounting, Finance and Banking, Business School, Manchester Metropolitan University, Oxford Road, M15 6BH, Manchester, United Kingdom E-

mail: b.banna@mmu.ac.uk

Early Warning System to Predict Energy Prices: The Role of Artificial Intelligence and Machine Learning

Abstract

The COVID-19 pandemic has inflicted the global economy and caused substantial financial losses. The energy sector was heavily affected and resulted in energy prices massively tumbling. The Russian invasion of Ukraine has fueled the energy market more volatile. In such uncertain contexts, an Early Warning System (EWS) would efficiently contribute to stabilizing market swings. It will leverage the ability to control operating costs and pave the way for smooth economic recovery. Within this framework, we deploy Machine Learning (ML) models to forecast energy equity prices by employing uncertainty indices as a proxy for predicting energy market volatility. We empirically examine the comparative effectiveness of prevalent ML models and conventional approaches (regression) to forecast the energy equity prices by utilizing the daily data from 1/6/2011 to 18/1/2022 for four US uncertainty and eight energy equity indices. Results show that the Nonlinear Autoregressive with External (Exogenous) parameters (NARX) of Neural Networks (NN) scored significantly better accuracy than all other (25) ML models and conventional approaches. The study outcomes are beneficial for policymakers, governments, market regulators, investors, hedge and mutual funds, and corporations. They improve stakeholders' resilience to exogenous shocks, blaze the recovery path, and provide evidence-based for assets allocation strategies.

JEL: Q47, C53, E17, C32, H12

Keywords: *Energy equity prices; Machine Learning; Early Warning Systems; Forecasting; COVID-19; United States.*

1. Introduction

In late 2019, the outbreak of the COVID-19 pandemic caused major disruptions with heightened uncertainty and lockdowns that led to increased unemployment rates and severe economic slowdown. Estimated as one of the most painful economic crises since the Great Depression (1929–1933), it resulted in massive financial and economic damages worldwide with far-reaching consequences on most industries (Alshater et al., 2021; Banna et al., 2022; Gupta et

al., 2020). The global energy system witnessed a significant drop in energy consumption with rising uncertainty about energy supply (Soava et al., 2021; Shamsudheen et al., 2022) and consumption patterns (Hallack & Weiss, 2020; Olubusoye et al., 2021). Compulsory lockdowns significantly constrained human mobility, and manufacturing lockdown damaged economic activities leading to a shrink in energy demand. This unprecedented shock caused a dramatic drop in energy equities prices and severe market instability. Remarkably, the US energy market, one of the largest global energy consumers, was dramatically hit. For the first time in history, the West Texas Intermediate (WTI) crude oil future prices crashed from USD 85 per barrel on January 15, 2020, to a negative price of USD 40.32 per barrel on April 20, 2020. The volatility of crude oil and other energy commodities resulted in devastating effects. They are considered the main production factors and lead strategic energy sources for the global economy (Ben Jabeur et al., 2021; Tien & Hung, 2022). While the market slowly recovered, the Russian invasion of Ukraine changed the market dramatically and resulted in a more volatile energy sector.

This economic turmoil raised investors' concerns worldwide, and an approach to mitigate its harmful consequences became crucial. The present paper aims to examine three essential aspects of the US energy equity markets. First, we predict energy prices through uncertainty indices. Several studies were among the first to examine the connectedness effects between uncertainty indices and market volatility; they found that uncertainty indices can strongly predict oil prices (Bekiros et al., 2016; Balciilar et al. 2017; Chen et al., 2021). Second, we compare conventional methods with machine learning methods as many studies found that the latter group outperforms the former group (Ghoddusi et al., 2019). Finally, we determine the most effective method for an Early Warning System (EWS) for the energy sector during uncertain economic conditions; for instance, Okur et al. (2021) stated that few studies conducted a comparative examination of superior Machine learning (ML) models during uncertain and extraordinary economic conditions.

We employ ML and conventional approaches to achieve the objective of this study. The ML is especially effective for issues that cannot be addressed directly by an analytical solution, model matching, or complex regression and classification tasks. ML models have gained popularity in many aspects of the energy business, inter alia energy prices. This is due to better data processing, categorization, proactive use of large-scale complex information and a large pool of complex data, as well as overlapping dynamics associated with a high level of uncertainty. In

this context, predicting energy prices requires more sophisticated models than conventional approaches (Castelli et al., 2020). ML is the most reliable approach as it allows to focus on nonlinear dependencies and exchanges between forecasters (Naumzik & Feuerriegel, 2021). It also improves forecasting accuracy using all possible input variables (Ghoddusi et al., 2019). Effectively, accurate prediction of the highly volatile energy prices constitutes a reliable reference for controlling the costs, grasping market trends and opportunities, and providing a scientific tool and evidence-based data for policymakers and market regulators (Lu et al., 2021). Accordingly, the integration of EWS in energy economics and finance becomes indispensable to achieve major goals like lowering operational expenses, securing energy sectors, and contemplating economic revival.

The contributions of this paper to the extant literature manifold are as follows. First, this study, to the best of our knowledge, tests the power of EWS to predict the energy prices during the pandemic. Second, it contributes to evaluate the risk of uncertainty during the COVID-19 period in the USA by using uncertainty indices as independent variables (predictors) and energy equity indices as dependent variables (responses). The analysis was performed in the context of USA since they are the largest global energy consumers, and related data is available. Third, it performs the predictability simulation by relying on conventional approaches (like regression analysis) and ML models during the pandemic to highlight ML superiority and provide accurate energy price forecasts. Fourth, it extends the ML literature by incorporating around 26 Artificial Intelligence (AI) predictive approaches and highlighting its effectiveness in uncovering key determinants for energy price predictability. Our results find that the Nonlinear Autoregressive with External (Exogenous) parameters (NARX) of Neural Networks (NN) scored significantly better accuracy than all other ML models and conventional approaches.

The remainder of the paper is structured as follows. Section 2 synthesizes the related literature review. Section 3 describes the data and methodology, while the empirical results are presented in Section 4. Section 5 highlights our discussions and Section 6 concludes the analysis.

2. Literature Review:

In the United States, the energy system is built from a massive interconnected network that produces and distributes energy from a wide range of energy sources to keep pace with the increasing demand. Industry dynamics are driven by a variety of complex, uncertain and diverse factors beyond supply and demand frameworks. Exogenous shocks such as the Middle East crises,

subprime mortgage financial crises, natural disasters, economic uncertainty, and other extreme events such as coronavirus pandemic influence energy stock prices and aggravate their volatilities (Ftiti et al., 2020; Ren et al., 2021; Scarciuffolo & Etienne, 2021). Understanding the complex linkages and dependencies between energy prices and external factors is therefore crucial for investors and policy makers to develop and evaluate appropriate strategies and alternatives in times of crisis. A growing literature strand paid attention to finding accurate predicting mechanisms for stock price volatility which has a significant impact on energy consumption. In doing so, many researchers employed conventional methods (regression, e.g., ARIMA, GARCH) to predict volatility. Yet, in the past decade, another approach emerged based on machine learning (ML) and Artificial Intelligence (AI).

ML is a data science-driven model that can identify existing data patterns and enhance temporal effectiveness. The beginning of ML dates back to the 1950s and 1960s, when researchers' curiosity led them to mimic human learning through computers. Accordingly, insightful information can be extracted, which can subsequently be utilized for the projection and generation of new knowledge. Initially, ML implementation started to flourish in the economics and finance fields (Ahmed et al., 2022). One of its earliest uses in energy economics was to forecast electricity prices. In the 2000s, ample publications attempted to estimate power costs using traditional Artificial Neural Network (ANN) methodologies. For instance, Khosravi et al. (2013) employed the delta and bootstrap methodologies to generate power price prediction intervals (PIs) for parameter uncertainty. Papadimitriou et al. (2014) evaluated the predictive power of Support Vector Machines (SVM)-based forecast models to assess the next-day direction shift in power prices. Their findings indicate that SVM is a solid strategy for forecasting short-term power prices with a predictive efficiency of 76.12% across a 200-day timeframe.

Another literature strand invigorated the use of ML and econometric models. This trendy combination gained popularity and attracted scholars' attention. Godarzi et al. (2014) created a dynamic Non-Linear Autoregressive model incorporating Exogenous inputs (NARX), similar to the Levenberg-Marquardt approach, using the Ensemble Empirical Mode Decomposition (EEMD) approach. Zhang et al. (2015) decomposed global crude oil prices into a variety of separate Intrinsic Mode Functions (IMFs) and residual terms. They also created the Least Square Support Vector Machines (LSSVM) - Particle Swarm Optimization (PSO) approach. They applied the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model to anticipate the time-varying and

non-linear components of crude oil prices. The former models frequently relied on an autoregressive framework. In contrast, the latest studies employed a hybrid strategy capable of managing a diverse collection of input factors such as need, supply, and equity market indices (Chai et al., 2018). Dogah and Premaratne (2018) investigated the susceptibility of BRICS stock returns to changes in oil risk indicators. They integrated the Vector Autoregressive (VAR) model with a random forest approach to address some VAR shortcomings and tackle oil-risk issues.

In this regard, it is essential to expose the benefits of ML sophisticated approaches and the attributes of its superiority over traditional economic models. The scientific community reported a marginally significant increase in predicting out-of-sample forecast accuracy and performance assessment measures. To illustrate the disparities between actual and anticipated values, a standard key metric, the Root Mean Square Error (RMSE), was used. A low RMSE implicates better predicting power. In the same vein, Collado and Creamer (2016) anticipated natural gas prices using an approximation dynamic programming technique that combines "a time series method (ARIMA) with two machine learning algorithms (Support Vector Machine and Random Forests). This technique surpasses logistic regression, which was treated as a baseline. Debnath and Mourshed (2018) explored the prediction model for Energy Planning Models (EPMs) and discovered that Artificial Neural Network (ANN) is the most used approach. Athey (2017) and Mullainathan and Spiess (2017) presented a non-technical concise summary and evaluation of machine learning's economic applicability. The economic sector recently allocated great importance to exploring ML capabilities and limits in testing hypotheses and causation, especially in the wake of the COVID-19 pandemic.

However, a more sophisticated variant of ANNs, the Deep Learning (DL) models, is still rarely used in predicting crude oil prices. Zhao et al. (2017) exceptionally applied the DL algorithm "Stacked Denoising Autoencoders (SDAE) with bootstrap aggregation in the research (bagging). The bagging process created a large number of data sets for training a series of SDAE base models. Tang et al. (2015a) presented an ensemble paradigm that combined Extended Extreme Learning Machine (EELM) and Complementary Ensemble Empirical Mode Decomposition (CEEMD) to improve petroleum prices' predictive accuracy. The findings revealed that the model is a potential prediction tool for complex time series data with significant instability and irregularities. Zhu et al. (2016) proposed an Adaptive Multiscale Ensemble Learning (AMEL) paradigm that integrated LSSVM, PSO, and EEMD with a kernel function prototype.

Other studies include Dudek (2016) who proposed a probabilistic power price forecasting strategy based on a Feed-Forward Neural Network (FFNN) precluding the pre-processing of extra data. Panapakidis and Dagoumas (2016) studied predicting models that include ANN-based prices for the day ahead. To anticipate energy prices, Wang et al. (2017) proposed a two-layer decomposition strategy before building a hybrid model built on Fast Ensemble Empirical Mode Decomposition (FEEMD), Variational Mode Decomposition (VMD), and BPNN and enhanced with a swarm intelligence-based approach. The model outperformed the one-step and multi-step power price estimates. To address the limitations of the classic ANN model, Singh et al. (2017) used a modified neuron model to anticipate the short-term power pricing of the energy market in Australia.

Table 1 presents the previous studies in the field and reports their contextual methodological use.

Table 1: Machine Learning use in energy economics and finance

Article info	Title
Jia et al. (2007)	SVM model with multiple inputs and single output for China's coal needs.
Xuemian and Guohao (2008); Yang et al. (2014); Yun-cai (2003)	Simple ML models to forecast coal usage.
Khosravi et al. (2013)	Create power prices Prediction Intervals (PIs) for uncertainty quantification through delta and bootstrap approaches.
Fang et al. (2013)	Artificial neural networks (ANN) to derive the quantitative coefficients of the underlying system for predicting carbon prices on fuel efficiency.
Perera et al. (2014)	ML applications for renewable energy integration.
Weron (2014)	ML for forecasting power prices.
Godarzi et al. (2014)	Develop a dynamic Non-Linear Autoregressive model with Exogenous inputs (NARX).
Papadimitriou et al. (2014)	Investigate the efficacy of SVM-based forecasting models for predicting the next-day directional changes in electricity prices.
Zhang et al. (2015)	Ensemble Empirical Mode Decomposition (EEMD) approach for forecasting global crude oil prices.

Tang et al. (2015a)	Combine Extended Extreme Learning Machine (EELM) and Complementary Ensemble Empirical Mode Decomposition (CEEMD) to improve petroleum prices predictive accuracy.
Collado and Creamer (2016)	The approximation dynamic programming technique combines a time series method (ARIMA) with two machine learning algorithms (Support Vector Machine and Random Forests) for forecasting natural gas prices.
Panapakidis and Dagoumas (2016)	ANN-based for forecasting the price for the day ahead.
Dudek (2016)	Probabilistic power price forecasting based on a Feed-Forward Neural Network (FFNN).
Zhu et al. (2016)	Adaptive Multiscale Ensemble Learning (AMEL) paradigm that integrates Least Square Support Vector Machines (LSSVM), PSO, and EEMD with a kernel function prototype.
Singh et al. (2017)	Generalized neuron model to forecast the short-term power price of the Australian energy market.
Wang et al. (2017)	Develop a hybrid model centered on Fast Ensemble Empirical Mode Decomposition (FEEMD), Variational Mode Decomposition (VMD), and BPNN improved using a swarm intelligence-based algorithm for energy prices.
Zhao et al. (2017)	Stacked Denoising Autoencoders (SDAE) — a DL approach for predictions related to crude oil prices.
Athey (2017); Mullainathan and Spiess (2017)	Non-technical overview and assessment of ML's economic/econometric applications.
Zemene and Khedkar (2017)	ML algorithms for calculating client-electric power use.
Voyant et al. (2017)	ML techniques for predicting solar radiation.
Debnath and Mourshed (2018)	Forecasting model for Energy Planning Models (EPMs) using Artificial Neural Network (ANN).
Dogah and Premaratne (2018)	Combine the VAR model with the Random Forest technique to forecast the sensitivity of stock returns in BRICS.
Farajzadeh and Nematollahi (2018)	WNNs to explore the performance of regression models in predicting energy intensity and its components.

3. Data and Methodology

3.1 Data

This study uses daily data between January 1, 2011 and January 18, 2022. The sample contains 2677 observations, and the data was collected from Thomson Reuters DataStream and the policy uncertainty website. To assess the predictive capability of ML models, we split the sample into two periods: the pre-pandemic period from 1/6/2011 to 31/12/2019 and the COVID19 pandemic period from 1/1/2020 to 18/1/2022¹. We spot three critical dates related to COVID19 events. The first date was January 27, 2020, when the World Health Organization (WHO) described the COVID -19 as a global threat based on assessed risk worldwide. The second date was February 24, 2020, marking two critical events: (1) a significant increase in COVID-19 cases outside mainland China and (2) the collapse of the global market. The third date was September 3, 2020, when a complete global lockdown was put in place. The indices are presented in Table 2, along with the study descriptive statistics. The data is separated into Panels A and B. Panel A includes the energy equity indices, the dependent variables (responses). Panel B contains the economic uncertainty indices, the independent variables (predictors). Precisely, Panel A consists of the US Renewable Energy Equity index, US Oil, Gas, and Coal Equity index, US Pipelines Equity index, US Oil Equity & Services index, US Oil & Gas Refining & Marketing Equity index, US Oil: Crude Production Equity index, US International Oil & Gas index, and the main US Energy index. In the same framework, Panel B consists of the Twitter Economic Uncertainty (TEU-USA) index, the EMV Infections Uncertainty (EMV) index, the Economic Policy Uncertainty (EPU) index, and the Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX), respectively. The description and sources of the variables are illustrated in Appendix 1.

In Panel A, almost all indices are left- or right-skewed since means are less than their respective medians and vice versa. Based on standard deviation, US Oil & Gas Refining & Marketing Equity index has the highest volatility while the US Renewable Energy Equity index

¹ The first cases of COVID-19 were recorded in the first half of January 2020 in the mainland of China. Thus, we consider the beginning of the pandemic period the date 1/1/2020.

and US pipelines have the lowest. In almost all cases, the kurtosis is less than 3, which means the indices have lighter tails than in a normal distribution. Furthermore, most indices are negatively skewed, which implies that extreme changes occur more frequently. In Panel B, TEU-USA and

the EPU index are the most volatile, while the VIX is the least. All indices are right-skewed and have kurtosis higher than three. On the other hand, skewness is positive for all variables. Furthermore, the Augmented Dickey-Fuller (ADF) tests for unit roots are accepted for all indices; the more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence. However, after computing the differences between consecutive observations, the dataset becomes stationary. This method is known as differencing (1-lag). The Phillips–Perron (PP) test confirms the ADF unit roots results.

Panel A: Energy Equity Indices

	<i>US Energy Eq</i>	<i>Renew US Oil, Gas, Coal</i>	<i>US Pipelines</i>	<i>US Oil Eq & Svs</i>	<i>US Oil Ref, Mkting</i>	<i>US Oil: Crude Prod.</i>	<i>US Int. Oil & Gas</i>	<i>US Energy</i>
Mean	273.40	2009.06	764.44	1194.59	49142.62	1024.45	2298.52	1843.81
Standard Error	3.654	7.555	3.820	9.178	290.242	5.396	6.637	7.994
Median	217.05	2036.41	741.10	1273.23	50427.66	1034.22	2340.49	1863.30
Standard Deviation	189.08	390.88	197.63	474.88	15017.03	279.21	343.40	413.59
Sample Variance	35749.9	152789.8	39057.6	225507.7	225511155.1	77958.0	117922.0	171059.3
Kurtosis	3.61	0.64	-0.29	-0.83	-0.14	0.04	1.09	0.16
Skewness	2.00	-0.60	0.21	-0.25	-0.07	-0.28	-0.91	-0.35
Range	1001.50	2188.03	979.03	2011.73	70396.29	1433.48	1989.20	2166.93
Minimum	47.55	740.97	249.13	210.79	16161.59	267.26	1052.46	630.83
Maximum	1049.05	2929.00	1228.16	2222.52	86557.88	1700.74	3041.66	2797.76
Sum	731882.9	5378244.4	2046413.9	3197927.5	131554784.8	2742463.8	6153127.8	4935872.6
Count	2677	2677	2677	2677	2677	2677	2677	2677

Panel B: Uncertainty Indices

	<i>TEU - USA</i>	<i>EMV Infections Uncertainty</i>	<i>US Economic policy uncertainty index</i>	<i>CBOE VIX Index</i>
Mean	101.90	3.36	119.46	17.73
Standard Error	2.064	0.152	1.744	0.142
Median	71.34	0.34	92.88	15.7
Standard Deviation	106.80	7.88	90.25	7.35
Sample Variance	11405.2	62.1	8144.3	54.0
Kurtosis	26.05	14.48	8.60	13.09
Skewness	3.94	3.44	2.48	2.80
Range	1557.63	68.37	804.34	73.55
Minimum	2.63	0	3.32	9.14
Maximum	1560.26	68.37	807.66	82.69

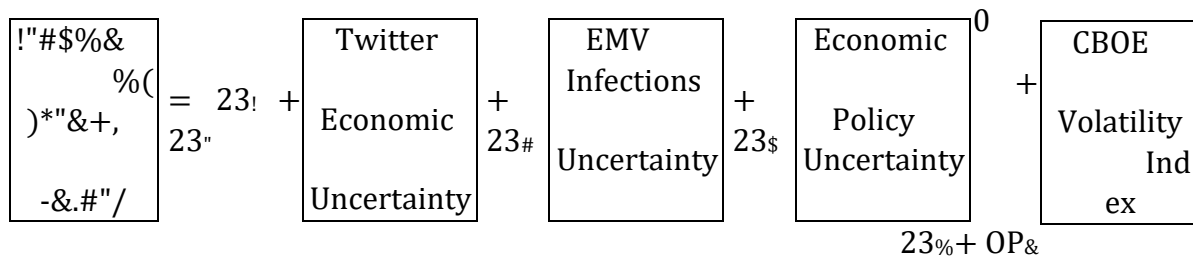
Sum	272795.0	8983.8	319794.1	47464.8
Count	2677	2677	2677	2677

Table 2. Descriptive Statistics

3.2 Methodology

This study uses ML and traditional approaches (i.e., regression) to forecast the prices of energy equity indices. The Root Mean Square Error (RMSE) is deployed to examine models' implementation alongside the accuracy of the predictability. In Table 1 above, we gathered previous studies and pinpointed respective methodologies in chronological order. Notably, previous studies have used one- or two-ML approach(es) in combination with conventional models (such as regression) to predict the results. In the present paper, we combine almost all available ML models in the literature with conventional models including Multiple Linear Regression model (MLR) to predict the energy prices pre-and during the COVID-19 pandemic. The last period is characterized by the high level of uncertainty. It thus serves to assess whether economic uncertainty indices can be helpful in measuring the impact of energy economics and finance during the pandemic. As previously mentioned, the demand for energy has dramatically fluctuated during the pandemic, making it difficult to predict the actual needs of supplies. Following the framework of this approach, the MLR is most suitable method to proceed with predictors and responses, as it dramatically fits our model structure and the set of dependent and independent variables (Aertsen, 2010; Naumzik & Feuerriegel, 2021; Raju & Laxmi, 2020).

The present study employed the following approach to predict the prices of the energy sectors:



For sectors of energy prices, we use the energy equity sector indices as indicated in Panel A of Table 2. First, we estimate the energy sectors using a simple Multiple Linear Regression model (MLR) and check the performance. Second, we evaluate the results of Supervised Regressionbased Machine Learning models and compare them to MLR to determine which strategy is more effective in forecasting energy prices. We run both models for both periods with the same features to cross-check models' performance (in-sample) and during the uncertain period of COVID-19 (out-of-sample).

We follow an exhaustive application of ML methods previously used in the literature (ML Generalized Linear Model (GLM), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Regression Trees, Ensemble Methods, and Neural Networks). The last method was gauged following the Deep Learning family and specifically the Levenberg-Marquardt algorithm for non-linear least squares modes to train networks with dynamic time series². The accuracy of the model is calculated with Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

4. Empirical Results

This paper extends the literature by combining various ML models to propose the most suitable technique for predicting volatility in times of crisis by comparing its efficiency with conventional methods. Table 3 presents the results of the Multiple Linear Regression analysis pre and during COVID-19. In the pre-pandemic period, R-Squared is significantly low in all cases. However, during COVID-19 most of the variables are statistically significant, which means that the uncertainty indices are suitable and can be used to predict the energy equity prices. F-Statistic is significant in all cases, while R-Squared varies from 0.44 to 0.67. Not surprisingly, the variables of the indices are negatively correlated with energy prices³. Lastly, the RMSE looks satisfactory in all cases except the US Oil & Gas Refining & Marketing Equity index. The RSME looks high in pre and during COVID-19 periods at a 13,798 and 6,534, respectively. Finally, RSME is lower in the COVID-19 pandemic as increased uncertainty explains the co-movements of energy prices and uncertainty indices.

The intercept parameter is positive and statistically significant in all cases. The mean of intercept is the value of the dependent variable when the independent variable equals zero. As the values of the parameters are relatively high for estimations, it can be considered that there are significant factors that affect the energy prices not related to the uncertainty indices. Moreover, the VIX index is negatively correlated in all estimations but not statistically significant in the case

² More information about methodological and technical parts of the machine learning models applied in the study can be found in the online appendix.

³ This is something that makes sense, since an increase in uncertainty would have a negative impact on investment activities, and as a result, this will lead to a decrease in energy prices.

of US Oil Equity & Services and US Oil & Gas Refining & Marketing in the pre-pandemic. At the same time, during the pandemic, it's not statistically significant in the case of US Renewable Energy index⁴. The EMV Infections Uncertainty index is not statistically significant except with US Pipelines Equity index (pre-pandemic) and US Renewable Energy index (during a pandemic). The US Economic Policy Uncertainty and US Twitter Economic Uncertainty indices are statistically significant in almost all cases with remarkably small values. Both indices have the highest impact on predicting the energy prices in our model.

In Figures 4 and 5, we illustrate the actual and predicted prices along with residuals based on MLR analysis¹. During the COVID-19 pandemic, the prices become very volatile compared to the previous period, especially after the WHO announcement about the spread of COVID-19 globally; prices have become very unstable with dramatic decreases and extreme volatility till today. Also, there is no doubt that the MLR model struggles to effectively capture the predictions of the prices over that period (from January 2020 and beyond). In addition, as we mentioned above in Table 3, the residuals are less volatile during the COVID-19 pandemic. This makes sense since increased uncertainty during the pandemic explains the co-movements of energy prices and

¹ For the visual illustration, we present only the actual and the predicted prices with non-stationary data for comparison with the applied machine learning models in this study.

uncertainty indices. Lastly, the residuals of the US Oil & Gas Refining & Marketing Equity index are highly volatile, indicated by high RMSE value.

⁴ We consider it statistically significant if at 5% level.

Multiple Linear Regression analysis results - Pre pandemic period

	USRenewEnergy	USOilGasCoal	USPipelines	USOilEqSvs	USOilRefMkting	USOilCrudeProd	USIntOilGas	USEnergy
(Intercept)	177.64 (0.0000)	2413.81 (0.000)	960.93 (0.000)	1391.80 (0.000)	72577.23 (0.000)	1.81E-41 (0.000)	2783.31 (0.000)	2204.31 (0.000)
CBOEVolatilityIndex	2.6477 (0.0000)	-13.6102 (0.000)	-5.4559 (0.000)	1.4636 (0.3108)	1.20E-74 (0.28943)	-3.0975 (0.00031)	-20.6384 (0.000)	-9.8061 (0.000)
EMVInfectionsUncertainty	0.15 (0.9353)	1.9752 (0.73446)	18.6784 (0.000)	-11.5794 (0.1621)	553.76 (0.1090)	-4.0685 (0.4087)	9.566 (0.0298)	6.680 (0.3161)
EconomicPolicyUncertaintyIndex	-0.1847 (0.0000)	0.3020 (0.0022)	0.007 (0.9105)	1.1515 (0.000)	-80.6889 (0.000)	0.4273 (0.000)	-0.0656 (0.3796)	0.4799 (0.000)
TwitterEconomicUncertainty	0.0283 (0.3046)	-0.9437 (0.000)	-0.6781 (0.000)	-2.0206 (0.000)	34.8046 (0.000)	-1.0045 (0.000)	-0.3294 (0.000)	-1.3166 (0.000)
RMSE	73.83	232.63	146.41	330.82	13798.96	196.72	175.85	266.20
R-squared	0.04	0.17	0.16	0.11	0.25	0.10	0.33	0.14
Adjusted R-Squared	0.04	0.17	0.16	0.11	0.25	0.10	0.33	0.14
F-statistic	22.1 (6.38e-18)	111 (6.46e-86)	104 (7.53e-81)	70 (8.77e-56)	183 (3.18e-135)	61.6 (2.37e-49)	261 (2.72e-183)	87.2 (9.64e-69)

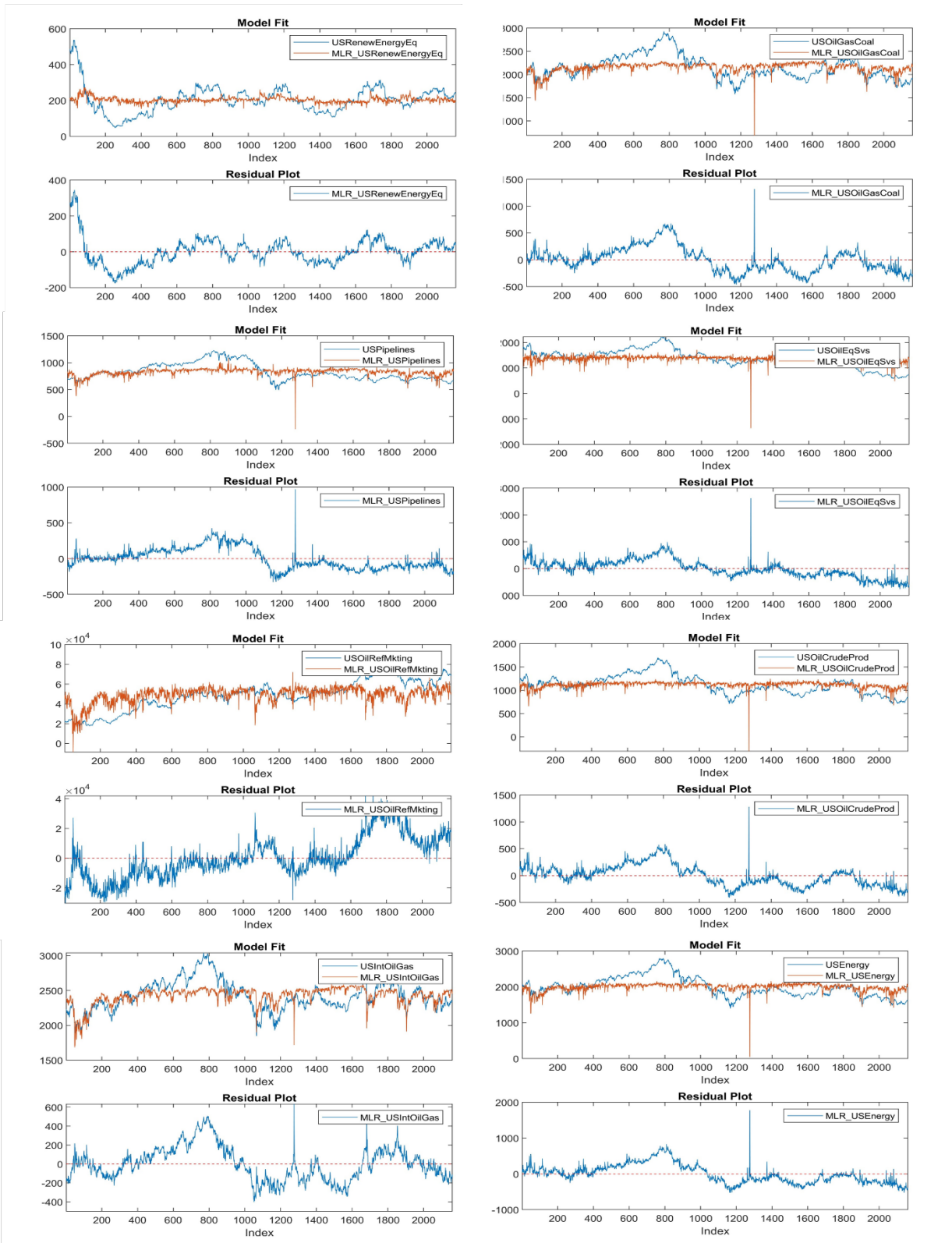
Multiple Linear Regression analysis results - COVID-19 pandemic period

	USRenewEnergy	USOilGasCoal	USPipelines	USOilEqSvs	USOilRefMkting	USOilCrudeProd	USIntOilGas	USEnergy
(Intercept)	687.99 (3.543e-89)	1847.08 (1.137e-246)	645.62 (1.154e-284)	660.72 (7.075e-241)	65135.47 (3.698e-249)	8.89E+02 (4.149e-180)	2210.74 (3.590e-254)	1586.73 (1.527e-255)
CBOEVolatilityIndex	3.0297 (0.0932)	-8.8758 (2.422e-06)	-3.3041 (1.714e-09)	-3.3967 (9.911e-07)	-3.96E+02 (2.013e-09)	-3.5913 (0.0050)	-10.1306 (3.009e-06)	-7.3012 (2.349e-06)
EMVInfectionsUncertainty	3.4512	-0.4063	-0.1775	0.0859	53.09	-0.6453	-0.3476	-0.1149

	(0.0034)	(0.7377)	(0.6130)	(0.8476)	(0.2092)	(0.4380)	(0.8036)	(0.9082)
EconomicPolicyUncertaintyIndex	-0.026 (0.8103)	-0.6946 (1.158e-09)	-0.2553 (2.029e-14)	-0.3216 (3.695e-14)	-22.0812 (2.517e-08)	-0.4353 (2.4105e-08)	-0.6192 (2.125e-06)	-0.5787 (6.8915e-10)
TwitterEconomicUncertainty	-1.2359 (2.216e-22)	-0.3177 (0.0113)	-0.0532 (0.1418)	-0.1771 (0.0001)	-11.9079 (0.0064)	-0.3225 (0.0001)	-0.2317 (0.1086)	-0.3323 (0.0013)
RMSE	181.66	187.65	54.30	69.13	6534.58	128.65	216.26	154.15
R-squared	0.44	0.58	0.64	0.67	0.59	0.59	0.49	0.61
Adjusted R-Squared	0.44	0.58	0.64	0.67	0.59	0.56	0.48	0.61
F-statistic	101 (1.53e-63)	180 (4.7e-96)	227 (7.16e-122)	261 (4.55e-122)	185 (6.29e-98)	162 (2.79e-89)	122 (7.05e-73)	201 (1.69e-103)

Notes: P-Values are in parenthesis.

Table 3. Multiple Linear Regression analysis results



Notes: The blue line presents the actual prices while the orange the predictions from the MLR model. In the upper right corner of every subfigure, we provide the name of each category of the energy sector. The depicted data is nonstationary.

Figure 4. Actual and Predicted prices with Residuals from the MLR model – Pre pandemic period

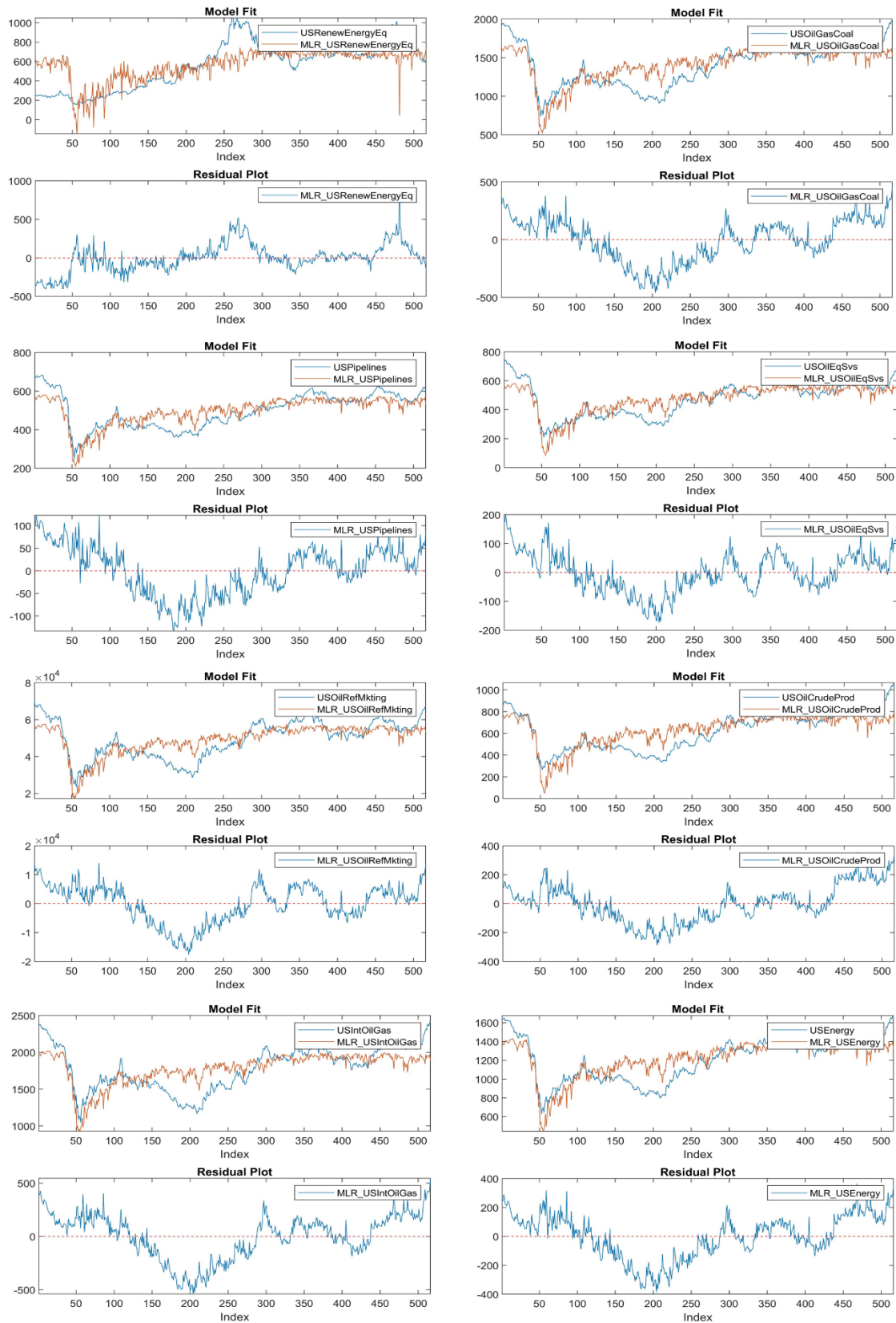


Figure 5. Actual and Predicted prices with Residuals from the MLR model – COVID-19 pandemic period

In Tables 4 and 5, we outline the training outcomes obtained out of ML models before and after the pandemic. Notably, ML models ultimately outperform the MLR approach. RMSE of most ML models is less than MLR, and R-Squared is higher under ML models. Figures 6 and 7 show the rankings of the predictions based on RMSE for all the employed models in this study. The Neural Network (NN) with the Levenberg-Marquardt algorithm clearly has the greatest result of all designs. The Gaussian Process Regression Models come second, while Regression Trees and SVMs come third and fourth. The Multiple Linear Regression model is always in the last position. In summary, ML models outperform conventional econometric approaches such as MLR.

However, it is worth mentioning the outstanding performance of forecasting of NN with the Levenberg-Marquardt algorithm. As shown in Figures 6 and 7, the NN has far less RMSE than any other model (machine learning and MLR). The key to this routine is hidden away in the NN structure. Figure 8 explains the structure of the NN with the Levenberg-Marquardt algorithm. The NN works as Nonlinear Autoregressive with Exogenous parameters. Specifically, $y(t)$ input is the actual price of the equity index, while the four exogenous input parameters are the uncertainty indices. In the model, we use ten hidden neurons and two delays (after optimization). The input summary includes an (observations) x (4 uncertainty indices) matrix, demonstrating complex data: (observations) timesteps of four elements (the uncertainty indices) while the output Targets an (observations) x 1 matrix, representing dynamic data: (observations) timesteps of one element (which is the predictions).

The network is formed and taught in open-loop mode, as illustrated in Figure 8. Closedloop training is less efficient than open-loop learning (multi-step). The open-loop allows us to supply correct historical results to the system while educating it to give accurate and current outcomes. Following training, the system may be turned into a closed-loop or really any design required by the app. There are three types of Timesteps: training, validation, and testing. During training, the network is supplied with timesteps, which is altered based on its inaccuracy. Validation timesteps are employed to evaluate network generalization and stop educating if adaptation is no longer improving. Because testing timesteps have no impact on learning, they give an objective audit of network quality both during and after learning. Prediction using these NN models is a type of dynamic filtering in which previous values with one or many time series are deployed to forecast subsequent values. Non-linear filtering and prediction are performed using dynamic neural

networks with tapped delay lines. Figures 9 and 10 present the time series response of the trained NN with the Levenberg-Marquardt algorithm. The residuals (verified through RMSE) are the least volatile of any other model employed in this study. The model looks pretty accurate during the pandemic in holding good persistence while predicting the actual prices. Lastly, regarding the validity of the models, Figure 8 presents the Mean Square Errors (MSE) of the most accurate model in the study, the Levenberg-Marquardt algorithm approach. The literature (e.g., Abedin et al., 2021; D'Ecclesia & Clementi, 2021; Yazici et al., 2020) only considers a welltrained model when the training and test sets are very similar. If the RMSE of the test set is much higher than that of the training set, the data is likely overfitted. In our study, the LevenbergMarquardt algorithm approach produces the best validation between 11 to 20 epochs, which is generally very low, proving the high accuracy of the model.

Overall, the Artificial Intelligence and Machine Learning models allow us to make more accurate predictions than the traditional Multiple Linear Regression models. The Neural Networks seem to be the most accurate from the Machine Learning models. This can be explained due to the flexibility to use autoregression simultaneously with exogenous input parameters, which makes the response outcome entirely accurate and minimizes RMSE relatively to the rest of ML models.

Model Type	USRenewEnergy			USOilGasCoal			USPipelines			USOilEqSvs		
	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared
Linear Regression												
Terms												
Linear	74.1	5487	0.03	234.1	54812	0.16	148.1	21934	0.14	336.4	113178	0.08
Interactions Linear	71.9	5167	0.09	227.5	51752	0.21	141.9	20136	0.21	314.5	98879	0.20
Robust Linear	76.9	5912	-0.04	234.2	54836	0.16	148.1	21942	0.14	337.9	114163	0.08
Stepwise Linear	71.8	5159	0.09	226.9	51493	0.21	141.7	20073	0.21	313.5	98289	0.2
Regression Tree												
Terms												
Fine Tree	83.3	6943	-0.22	249.6	62285	0.04	151.5	22940	0.10	351.1	123285	0.00
Medium Tree	75.7	5725	-0.01	224.1	50221	0.23	138.0	19050	0.25	319.4	102016	0.17
Coarse Tree	73.7	5429	0.04	215.8	46561	0.28	130.6	17064	0.33	307.8	94765	0.23
Support Vector Machines												
Terms												
Linear SVM	75.7	5734	-0.01	236.8	56074	0.14	153.1	23446	0.08	337.2	113711	0.08
Quadratic SVM	74.4	5529	0.03	255.6	65336	0.00	248.6	61807	-1.42	357.9	128064	-0.04
Cubic SVM	823.8	678663	-	577.6	333576	-4.12	1370.0	1876900	-72.5	734.0	538712	-3.36
			118.6									
Fine Gaussian SVM	73.5	5395	0.05	216.7	46968	0.28	135.0	18217	0.29	307.8	94735	0.23
Medium Gaussian SVM	71.5	5119	0.10	213.3	45488	0.30	128.9	16602	0.35	301.8	91095	0.26
Coarse Gaussian SVM	71.9	5176	0.09	225.5	50832	0.22	141.9	20141	0.21	313.9	98514	0.20
GPR Models												
Terms												
Squared Exponential GPR	70.5	4970	0.12	210.4	44277	0.32	127.7	16305	0.36	296.6	87966	0.29
Matern 5/2 GPR	70.0	4901	0.14	208.5	43460	0.33	126.7	16043	0.37	293.1	85902	0.30
Exponential GPR	69.3	4807	0.15	206.5	42651	0.34	125.6	15778	0.38	289.4	83735	0.32
Rational Quadratic GPR	69.4	4813	0.15	206.6	42688	0.34	126.0	15863	0.38	289.8	83961	0.32
Ensemble of Trees												
Terms												
Boosted Trees	71.2	5064	0.11	228.4	52157	0.20	130.7	17072	0.33	298.3	89001	0.28
Bagged Trees	70.3	4937	0.13	208.7	43535	0.33	125.6	15768	0.38	293.9	86389	0.30
Neural Networks												
Terms												
Narrow Neural Network	72.1	5202	0.08	232.1	53870	0.17	134.5	18085	0.29	309.9	96050	0.22
Medium Neural Network	72.3	5233	0.08	223.4	49916	0.23	133.6	17836	0.30	306.4	93863	0.24
Wide Neural Network	75.3	5666	0.00	220.5	48625	0.25	131.1	17190	0.33	325.2	105723	0.14
Bilayered Neural Network	72.2	5216	0.08	224.9	50580	0.22	131.8	17369	0.32	311.7	97138	0.21
Trilayered Neural Network	74.3	5515	0.03	213.1	45420	0.30	133.4	17785	0.30	298.7	89216	0.28
Levenberg-Marquardt alg.	24.4	593	0.96	31.1	969	0.99	26.3	691	0.99	30.0	901	0.99

Notes: We considered the Principal Components Analysis to transform features and remove redundant dimensions. Subsequent to training, a total of four components were retained. The explained variance noted for each component was as follows: 88.0%, 11.6%, 0.3%, 0.1%, respectively, for all ML models. For the Stepwise Linear Regression, the Maximum number of steps is one thousand. For the Fine, Medium, Coarse Trees, the Minimum leaf size was noted as (in order) 4, 12, and 36. For the Fine, Medium, and Coarse Gaussian SVM, the Kernel scale is 0.5, 2, and 8, respectively. For the Boosted and Bagged Trees, the Number of

learners is thirty and the Minimum leaf size eight. For the Narrow, Medium, Wide NN, the First layer size is 10, 25, and 100. For the Bilayered and Trilayered NN, the First- and Second-layer size is ten, and for the Trilayered NN, the Third layer size is ten. The iteration limit for all NN is one thousand. For the LevenbergMarquardt NN, we use ten hidden neurons and two delays.

Table 4. Machine Learning Training Results – Pre pandemic period (1/2)

Model Type	USOilRefMkting			USOilCrudeProd			USIntOilGas			USEnergy		
	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared
Linear Regression												
<u>Terms</u>												
Linear	13826	1.91E+08	0.25	198.6	39442	0.08	176.2	31036	0.32	268.4	72017	0.12
Interactions Linear	13796	1.90E+08	0.25	190.5	36298	0.16	173.7	30186	0.34	256.8	65957	0.20
Robust Linear	13874	1.92E+08	0.24	198.9	39569	0.08	176.1	31015	0.32	268.9	72329	0.12
Stepwise Linear	13810	1.91E+08	0.25	190.5	36298	0.16	174.4	30412	0.34	257.0	66059	0.20
Regression Tree												
<u>Terms</u>												
Fine Tree	15482	2.40E+08	0.06	211.0	44538	-0.03	195.2	38087	0.17	282.8	79948	0.03
Medium Tree	14095	1.99E+08	0.22	186.3	34693	0.19	180.3	32508	0.29	252.4	63681	0.23
Coarse Tree	13722	1.88E+08	0.26	180.8	32703	0.24	170.3	29009	0.37	244.3	59697	0.27
Support Vector Machines												
<u>Terms</u>												
Linear SVM	13976	1.95E+08	0.23	200.3	40120	0.07	177.6	31552	0.31	272.7	74343	0.10
Quadratic SVM	15502	2.40E+08	0.06	244.5	59780	-0.39	180.0	32404	0.29	330.0	108900	-0.32
Cubic SVM	92897	8.63E+09	-32.9	723.3	523148	-11.1	296.4	87841	-0.92	1451.3	2106272	-24.6
Fine Gaussian SVM	14095	1.99E+08	0.22	177.8	31627	0.27	174.4	30405	0.34	243.5	59287	0.28
Medium Gaussian SVM	13419	1.80E+08	0.29	178.0	31698	0.26	167.5	28060	0.39	240.2	57715	0.30
Coarse Gaussian SVM	13723	1.88E+08	0.26	187.7	35243	0.18	174.0	30283	0.34	256.0	65510	0.20
GPR Models												
<u>Terms</u>												
Squared Exponential GPR	13156	1.73E+08	0.32	173.6	30120	0.30	165.7	27453	0.40	234.4	54948	0.33
Matern 5/2 GPR	13096	1.72E+08	0.33	172.3	29677	0.31	164.2	26971	0.41	232.1	53856	0.34
Exponential GPR	13037	1.70E+08	0.33	171.0	29255	0.32	161.7	26147	0.43	229.9	52840	0.36
Rational Quadratic GPR	13052	1.70E+08	0.33	171.0	29255	0.32	162.2	26319	0.43	230.1	52946	0.36
Ensemble of Trees												
<u>Terms</u>												
Boosted Trees	13332	1.78E+08	0.30	179.7	32278	0.25	193.8	37539	0.18	248.7	61832	0.25
Bagged Trees	13125	1.72E+08	0.32	172.2	29656	0.31	165.6	27413	0.40	233.4	54480	0.34
Neural Networks												
<u>Terms</u>												
Narrow Neural Network	13830	1.91E+08	0.25	187.6	35201	0.18	171.7	29491	0.36	245.8	60393	0.27
Medium Neural Network	13573	1.84E+08	0.28	185.1	34277	0.20	172.3	29698	0.35	250.2	62590	0.24
Wide Neural Network	14443	2.09E+08	0.18	192.3	36995	0.14	166.8	27809	0.39	244.4	59741	0.27
Bilayered Neural Network	13580	1.84E+08	0.28	184.8	34133	0.21	167.5	28040	0.39	249.1	62026	0.25
Trilayered Neural Network	13483	1.82E+08	0.29	182.0	33120	0.23	171.8	29526	0.36	243.1	59083	0.28
Levenberg-Marquardt alg.	385	148235	0.99	29.0	840	0.99	43.5	1890	0.99	36.3	1319	0.99

Table 4. Machine Learning Training Results – Pre pandemic period (continued - 2/2)

Model Type	USRenewEnergy			USOilGasCoal			USPipelines			USOilEqSvs		
	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared
Linear Regression												
<u>Terms</u>												
Linear	183.2	33577	0.43	187.6	35183	0.58	54.6	2976	0.63	69.6	4843	0.66
Interactions Linear	183.1	33522	0.43	157.1	24665	0.71	45.3	2053	0.75	56.3	3172	0.78
Robust Linear	182.6	33357	0.43	188.0	35333	0.58	54.6	2980	0.63	69.7	4859	0.66
Stepwise Linear	182.5	33299	0.43	159.3	25367	0.7	45.6	2075	0.74	57.7	3328	0.77
Regression Tree												
<u>Terms</u>												
Fine Tree	182.8	33423	0.43	144.6	20895	0.75	43.4	1886	0.77	56.8	3225	0.78
Medium Tree	165.8	27473	0.53	138.3	19130	0.77	40.8	1663	0.8	54.3	2943	0.8
Coarse Tree	165.0	27222	0.54	151.0	22792	0.73	43.7	1910	0.76	55.8	3112	0.78
Support Vector Machines												
<u>Terms</u>												
Linear SVM	182.0	33131	0.44	191.6	36718	0.56	55.1	3033	0.63	69.9	4893	0.66
Quadratic SVM	175.5	30804	0.48	151.6	22980	0.73	44.8	2005	0.75	52.4	2745	0.81
Cubic SVM	210.8	44420	0.25	148.3	21999	0.74	41.6	1728	0.8	55.5	3081	0.79
Fine Gaussian SVM	166.0	27546	0.53	162.2	26322	0.69	47.6	2266	0.72	61.8	3817	0.74
Medium Gaussian SVM	152.5	23256	0.60	128.5	16517	0.80	37.3	1388	0.83	47.2	2224	0.85
Coarse Gaussian SVM	178.1	31709	0.46	156.1	24367	0.71	45.6	2075	0.74	56.4	3186	0.78
GPR Models												
<u>Terms</u>												
Squared Exponential GPR	149.6	22383	0.62	127.0	16132	0.81	36.3	1319	0.84	46.1	2124	0.85
Matern 5/2 GPR	145.6	21194	0.64	124.5	15508	0.82	35.5	1262	0.84	44.9	2016	0.86
Exponential GPR	141.7	20087	0.66	123.4	15220	0.82	35.1	1229	0.85	44.4	1970	0.86
Rational Quadratic GPR	142.8	20389	0.65	123.6	15277	0.82	35.4	1252	0.85	44.4	1972	0.86
Ensemble of Trees												
<u>Terms</u>												
Boosted Trees	153.2	23467	0.60	136.8	18703	0.78	41.7	1739	0.79	52.6	2766	0.81
Bagged Trees	147.8	21836	0.63	125.9	15861	0.81	36.2	1311	0.84	46.5	2164	0.85
Neural Networks												
<u>Terms</u>												
Narrow Neural Network	167.5	28050	0.52	136.3	18564	0.78	39.5	1560	0.81	49.7	2467	0.83
Medium Neural Network	154.1	23744	0.60	137.1	18802	0.78	42.1	1770	0.78	49.5	2450	0.83
Wide Neural Network	212.1	44982	0.24	168.2	28281	0.66	57.7	3334	0.59	55.2	3043	0.79
Bilayered Neural Network	159.3	25383	0.57	136.2	18556	0.78	37.0	1372	0.83	47.4	2247	0.84
Trilayered Neural Network	154.9	23994	0.59	130.1	16923	0.80	38.5	1486	0.82	47.2	2230	0.85
Levenberg-Marquardt alg.	43.3	1872	0.98	43.6	1901	0.99	34.7	1201	0.94	39.8	1583	0.97

Table 5. Machine Learning Training Results – COVID-19 pandemic period (1/2)

	USOilRefMktg	USOilCrudeProd	USIntOilGas	USEnergy
--	--------------	----------------	-------------	----------

Model Type	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared	RMSE	MSE	R-Squared
Linear Regression												
Terms												
Linear	6583.8	4.33E+07	0.58	129.9	16879	0.55	217.2	47185	0.48	155.8	24271	0.60
Interactions Linear	5658.6	3.20E+07	0.69	108.3	11738	0.69	185.8	34533	0.62	129.0	16641	0.73
Robust Linear	6587.4	4.34E+07	0.58	130.1	16929	0.55	217.7	47406	0.48	156.0	24330	0.60
Stepwise Linear	5806.3	3.37E+07	0.67	109.8	12058	0.68	186.8	34902	0.61	131.0	17171	0.72
Regression Tree												
Terms												
Fine Tree	5451.4	2.97E+07	0.71	102.1	10418	0.72	173.0	29919	0.67	118.9	14132	0.77
Medium Tree	5177.1	2.68E+07	0.74	98.8	9767	0.74	161.7	26134	0.71	110.0	12107	0.80
Coarse Tree	5309.2	2.82E+07	0.73	103.4	10681	0.71	172.8	29843	0.67	125.9	15848	0.74
Support Vector Machines												
Terms												
Linear SVM	6653.6	4.43E+07	0.57	129.7	16809	0.55	222.2	49391	0.46	157.7	24876	0.59
Quadratic SVM	5629.6	3.17E+07	0.69	105.6	11147	0.7	184.7	34110	0.62	120.0	14398	0.76
Cubic SVM	6374.8	4.06E+07	0.6	115.8	13405	0.6	192.4	37010	0.59	123.6	15284	0.75
Fine Gaussian SVM	5877.6	3.45E+07	0.67	112.7	12706	0.66	176.3	31089	0.66	139.4	19421	0.68
Medium Gaussian SVM	4791.4	2.30E+07	0.78	94.5	8922	0.76	155.1	24065	0.73	107.5	11554	0.81
Coarse Gaussian SVM	5748.0	3.30E+07	0.68	108.1	11690	0.69	186.6	34816	0.62	128.2	16440	0.73
GPR Models												
Terms												
Squared Exponential GPR	4749.8	2.26E+07	0.78	92.5	8553	0.77	153.3	23501	0.74	105.5	11132	0.82
Matern 5/2 GPR	4631.0	2.14E+07	0.79	90.2	8128	0.78	149.8	22443	0.75	103.6	10725	0.82
Exponential GPR	4527.6	2.05E+07	0.8	89.1	7935	0.79	147.0	21600	0.76	103.7	10752	0.82
Rational Quadratic GPR	4567.9	2.09E+07	0.8	89.1	7938	0.79	148.0	21904	0.76	103.1	10636	0.83
Ensemble of Trees												
Terms												
Boosted Trees	5091.8	2.59E+07	0.75	94.1	8851	0.76	169.0	28571	0.68	113.6	12898	0.79
Bagged Trees	4703.2	2.21E+07	0.79	90.3	8157	0.78	149.8	22443	0.75	103.1	10628	0.83
Neural Networks												
Terms												
Narrow Neural Network	4848.5	2.35E+07	0.77	94.8	8983	0.76	160.6	25796	0.72	109.9	12085	0.80
Medium Neural Network	5153.0	2.66E+07	0.74	108.9	11853	0.68	165.6	27413	0.7	111.4	12419	0.80
Wide Neural Network	7480.1	5.60E+07	0.46	143.6	20627	0.45	204.3	41738	0.54	141.8	20093	0.67
Bilayered Neural Network	4810.0	2.31E+07	0.78	118.9	14142	0.62	158.3	25062	0.72	111.6	12455	0.80
Trilayered Neural Network	5080.2	2.58E+07	0.75	96.7	9342	0.75	149.5	22359	0.75	120.9	14619	0.76
Levenberg-Marquardt alg.	679.0	461018	0.99	40.3	1624	0.98	51.6	2660	0.99	40.3	1624	0.99

Table 5. Machine Learning Training Results – COVID-19 pandemic period (continued - 2/2)

5. Discussions:

ML models have grown in popularity in many aspects of the energy industry due to their superior performance in processing, categorizing, as well as projecting complex and large-scale data. Comparing ML's features with classic econometric models (e.g., MLR) validates ML's rising appeal in energy economics analyses. The advantage of ML approaches stands in the ability of algorithms to handle massive volumes of structured and unstructured information while making quick decisions or forecasts. In this context, many studies relied on ML techniques, and some of them concluded to conflicting results. Some scholars have particularly applied ML in predicting solar radiation (Voyant et al., 2017), renewable energy integration (Perera et al., 2014), and calculating client-electric power use (Zemene and Khedkar, 2017). Weron (2014) explored ways for forecasting power prices and devoted a significant amount of the text to ML methodologies under the title Computational Intelligence (CI). Debnath and Mourshed (2018) investigated forecasting model for Energy Planning Models (EPMs) and found that Artificial Neural Network (ANN) is the most often utilized forecasting technique. Athey (2017) and Mullainathan and Spiess (2017) provided a non-technical overview and assessment of ML's economic/econometric applications.

Moreover, a methodology required a combination of ML and econometric models. For instance, Godarzi et al. (2014) developed a dynamic Non-Linear Autoregressive model with Exogenous inputs (NARX), similar to what we achieved using the Levenberg-Marquardt approach. Remarkably, most previous models have frequently employed a regressive framework, while recent studies used a hybrid strategy capable of managing a diverse collection of input factors such as need, supply, and equity market indices (Chai et al., 2018). Khosravi et al. (2013) created power prices Prediction Intervals (PIs) for uncertainty quantification through the use of the delta and bootstrap approaches, while Papadimitriou et al. (2014) investigated the efficacy of SVM-based forecasting models for predicting the next-day directional change in electricity prices. The literature has reported a marginally significant increase in predicting out-of-sample ability. Previous studies have relied on common performance indicators such as Root Mean Square Error (RMSE) to illustrate the disparities between actual and anticipated values. However, no study has conducted a thorough comparison and ranking of ML accuracy models as done in the present paper.

ML models have several advantages in dealing with data, processing, and analysis. First, ML can manage a diverse and vast number of inputs where the model's designer has little burden, particularly DL models, to select a minimal number of appropriate input variables. Secondly, the ML technique can accept dozens of potential input variables and choose the right components (or features) for prediction without respect to co-linearity problems. Third, ML algorithms are frequently capable of handling a wide range of theoretical and accurate data. This feature is particularly useful in the energy industry since it can combine text from articles or publications with time-series data to enhance projections. Fourth, ML models can also uncover complicated linkages and investigate various topologies for potential links across input and output data. Bayesian Model Averaging (BMA) methods make it possible to run a model with a specific data set in the space of classical models. Yet, this latter technique needs the modeler to specify each model's structure while ML models do not need a predefined structure. Fifth, they surpassed linear correlations to discover complicated, non-linear, and high-dimensional interactions among many input factors and intended outputs. Sixth, ML models are less sensitive to data quality. Traditional economic models have long been grappled with data scarcity difficulties (e.g., data sets that include outliers). Some methods (for example, Fuzzy and GNN models) were successfully developed within the ML field to provide the system with less worthy input. For instance, Alobaidi et al. (2018) provided guidance on ensemble models' capacity to deliver enhanced forecast performance with limited inputs. Seventh, ML models can function with little data pre-processing. Seasonal fluctuations, structure break, regime shifting, unit-root, and heteroskedasticity are all recognized properties of energy time-series data that should be tackled prior to implementing the econometric model. ML models do not require extensive data pre-processing since they may incorporate extra data features. Naturally, when features are customized to the demands of each projection via suitable ratios or kernel tweaks, ML algorithm performance would improve. In a nutshell, most ML approaches include advanced features to collect critical traits for the eventual prediction models.

Moreover, ML's primary focus is related to forecasting market values. Many noneconomic ML experts fail to recognize the core ML difference related to traded assets and physical events predictions. Predicting the temperature does not affect climate behavior, yet predicting crude oil and equity prices could immediately impact current prices and stimulate trade activities. In this regard, each ML algorithm that attempts to forecast prices using other algorithms creates an

externality. Each effective forecasting model makes the following algorithm more difficult and increases the chances of re-using underused approaches. A broad ML research topic is theorydriven which combines both theory and technique. For example, Gu et al. (2020) investigated several methods and discovered their superiority to depict commodity price changes, especially risk premia evaluation. The recent advances in Deep Learning approaches effectively transform the whole ML domain. Due to the multi-layer structure, DL techniques allow the algorithms to handle a substantially greater proportion of inputs in a very consistent manner without necessitating any prior feature specification. On the opposing hand, ML approaches often require a high volume of input data that requires additional effort and time.

The organizational and financial implications of machine learning and artificial intelligence in the energy business and other economic areas are constantly affected by ML innovations. The protracted impact on the energy industry's architecture has to be observed and analyzed and might replace or supplement human talents. In the replacement scenario, the automation will increase and replace human interventions, while ML is supposed to improve personnel' abilities in the supplementary role. As a result, an industry reorganization and new industry structures and actors are expected. Accordingly, future research should investigate and assess ML influence on critical aspects such as energy efficiency and cost, smart networks, the effectiveness of energy markets and exchanges, and the workforce in business during the pandemic. As such, the environmental disruption, pervasiveness of sustainable power, and spread of smart grids, machine learning, and artificial intelligence (ML/AI) could be utilized to augment forecasting abilities and integrate volatile renewable energy resources. This will contribute to balancing energy grids and understanding trends related to needs and demands.

Finally, though large energy commodities markets are very efficient, the advantages of price projection following advanced methodologies are still modest. Such benefits may become lower when giant firms adopt the technology to persuade the industry about its effectiveness while using latent data to reach the equilibrium prices, which is a serious hurdle effectively. Hence, ML becomes more useful in forecasting and anticipating market dangers by depicting the dispersion of shocks on equilibrium variables paired with network models and ML optimizing risk management strategies.

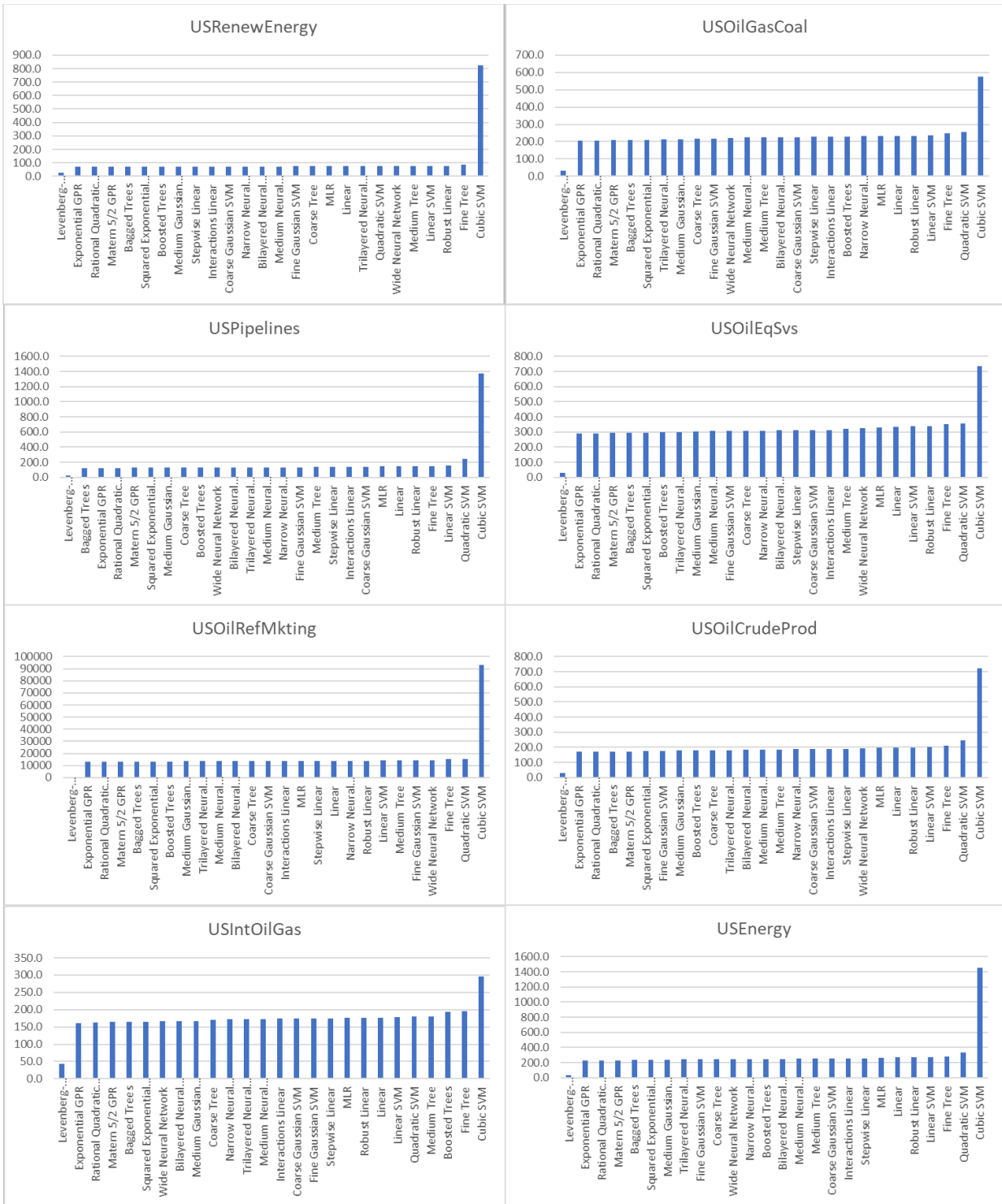


Figure 6. Rankings of RMSE for all forecasting models – Pre pandemic period

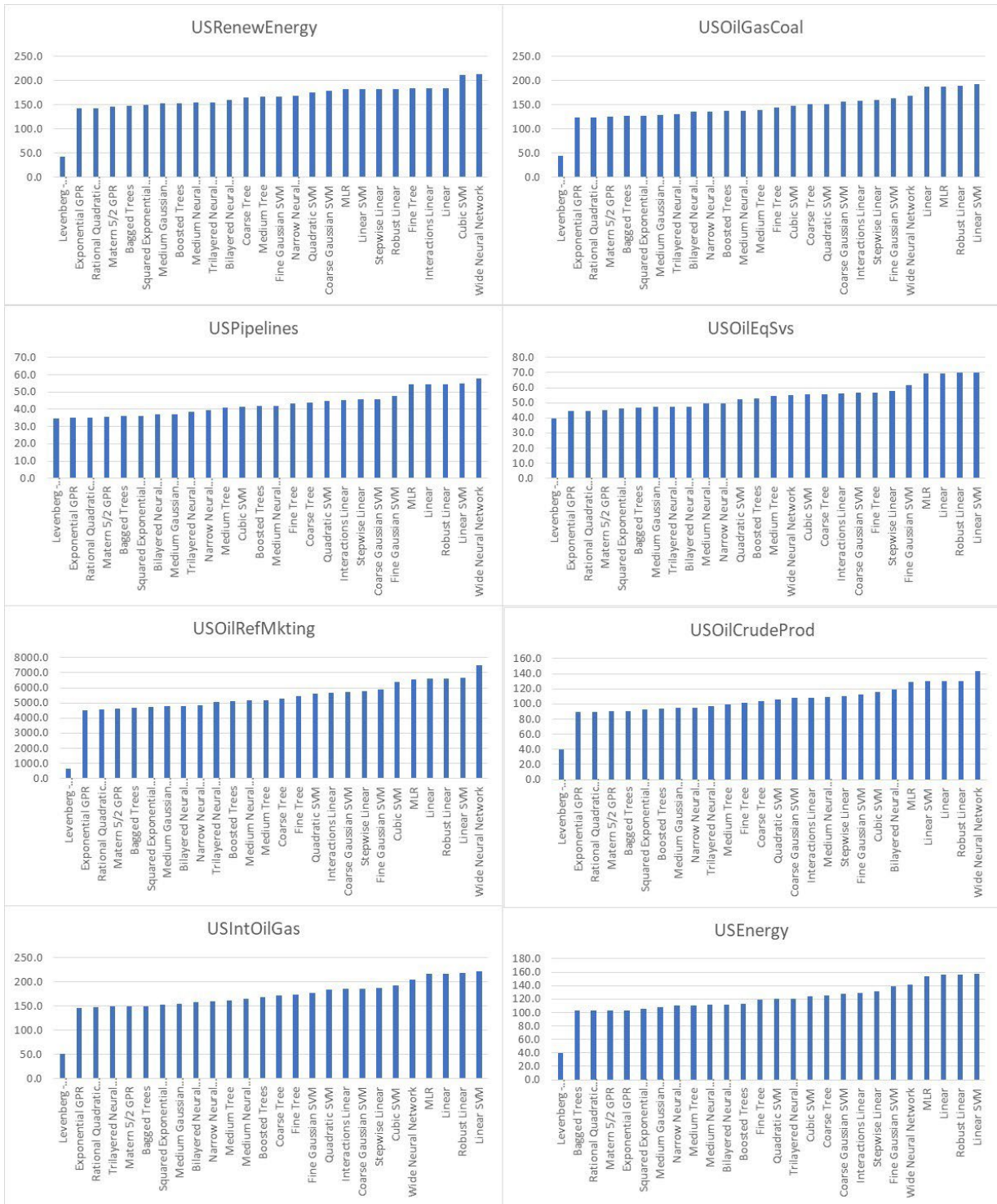
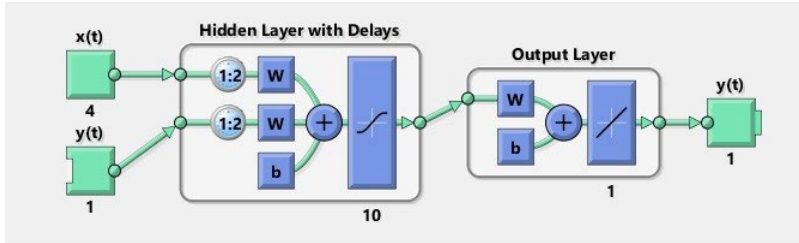


Figure 7. Rankings of RMSE for all forecasting models – COVID-19 pandemic period



Notes: We use ten hidden neurons and two delays. Also, in the upper left corner, we can see 4 exogenous parameters (the uncertainty indices), while in the bottom left, the energy prices as a Nonlinear Autoregressive with External (Exogenous) parameters (NARX) model.

Figure 8. Structure of Neural Network with Levenberg-Marquardt

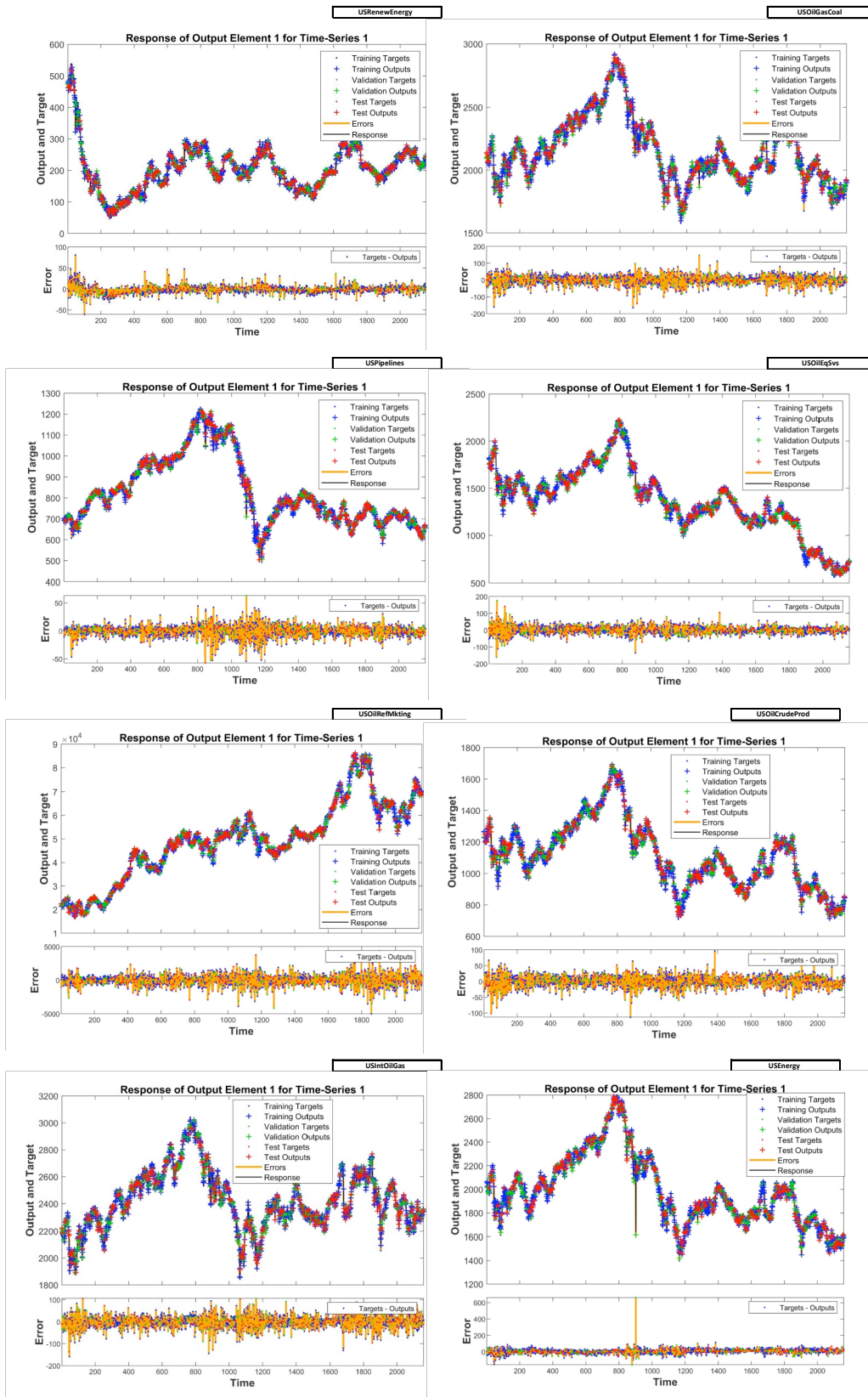


Figure 9. Time series Response of the trained N.N. with Levenberg-Marquardt algorithm – Pre pandemic period

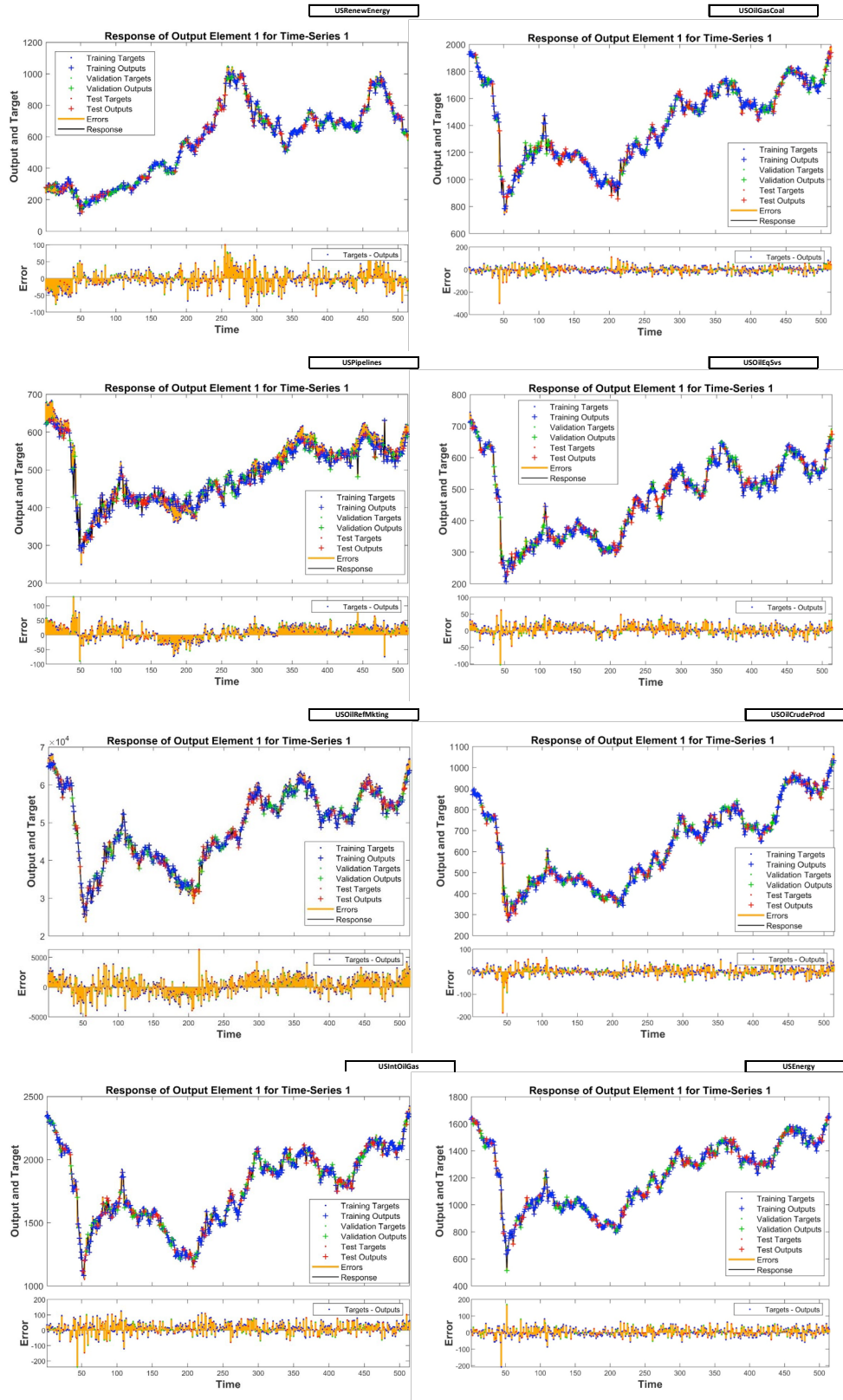


Figure 10. Time series Response of the trained N.N. with Levenberg-Marquardt algorithm – COVID-19 pandemic period

6. Conclusions

In the present study, we attempted to introduce an Early Warning System (EWS) to forecast energy equity prices and performed a comparative analysis based on ML and conventional regression models. We relied on equity indices since the valuation of such assets is tied up to expected corporate profits, assumed to positively affect investors demand. We gathered daily data from 1/6/2011 till 18/1/2022 to assess the predictability of the models before and during the COVID-19 pandemic. We used indices from renewable energy, oil, gas, coal, pipelines, and gas refining to proxy energy prices. We investigated the impact of economic uncertainty indices (Twitter Economic Uncertainty (TEU-USA); Infections Uncertainty (EMV); Economic Policy Uncertainty (EPU) index; and CBOE Market Volatility Index (VIX)) on the prediction of energy prices using RMSE as a measure of performance and accuracy.

We applied 25 ML approaches, including NN SVRs, Regression Trees, and GPR models. We found that ML models outperformed the MLR approach in all cases. The Nonlinear Autoregressive with External (Exogenous) parameters (NARX) was superior as it significantly improved accuracy. The simultaneous use of the Levenberg-Marquardt algorithm and autoregression with exogenous input parameters made the response outcome entirely accurate and minimized RMSE relatively to other NN machine learning models. In conclusion, the Artificial Intelligence (AI) and Machine Learning (ML) models allowed more accurate predictions than traditional Multiple Linear Regression models. The Neural Networks appeared the superior model.

Although the study contributed to important findings, it suffered from some limitations. First, the study was limited to the US context though it would be beneficial to perform it on global echelons. This is due to the availability of US data regarding the energy equity prices and uncertainty indices. Second, the high-frequency data cannot be easily accessed. In this regard, realtime data analysis might conclude valuable insights and important conclusions. Most databases provide high-frequency data for 3-month only, which restrict the analysis to daily frequency while forgoing the option to account for more extended period (two years or more), which can lead to valuable results.

We recommend expanding our suggested framework and the Early Warning Systems (EWS) and incorporating other sectors. While large energy commodities markets are exceptionally

efficient, price projection models based on advanced methodologies might conclude to undesirable outcomes as some powerful companies might benefit from this innovation and induce more effectiveness through improved latent data integration in the equilibrium prices. This can be considered a genuine jump based on sophisticated ML forecasts in the spectrum of energy economics. Machine learning appears valuable in determining and anticipating market threats when integrated with network architectures. It might improve the efficacy of risk management systems by accurately capturing the dispersion of disturbances on equilibrium parameters.

Competing interest statement:

The authors declare that there is no conflict of interest.

Acknowledgment:

The authors equally contributed to this paper.

References

- Abedin, M. Z., Moon, M. H., Hassan, M. K., & Hajek, P. (2021). Deep learning-based exchange rate prediction during the COVID-19 pandemic. *Annals of Operations Research*, 1-52.
- Abdella, M., Marwala, T., 2005. The use of genetic algorithms and neural networks to approximate missing data in database. *Computational Cybernetics*, 2005. ICCCYB, 2005. IEEE 3rd International Conference on. IEEE., pp. 207–212. <https://doi.org/10.1109/ICCCYB.2005.1511574>.
- Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K., & Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological modelling*, 221(8), 1119-1130.
- Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. *International Journal of Electrical Power & Energy Systems*, 31 (1), 13–22.
- Alshater, M. M., Atayah, O. F., & Khan, A. (2021). What do we know about business and economics research during COVID-19: a bibliometric review. *Economic Research Ekonomiska Istraživanja*, 1–29. <https://doi.org/10.1080/1331677X.2021.1927786>
- Akintande, O. J., & Olubusoye, O. E. (2020). Datasets on how misinformation promotes Immune perception of COVID-19 pandemic in Africa. *Data in Brief*, 31, Article 106031.
- Akintande, O. J., Olubusoye, O. E., Adenikinju, A. F., & Olanrewaju, B. T. (2020). Modelling the determinants of renewable energy consumption: Evidence from the five most populous nations in Africa. *Energy*, 206, Article 117992.
- Alobaidi, M.H., Chebana, F., Meguid, M.A., 2018. Robust ensemble learning framework for dayahead forecasting of household based energy consumption. *Applied Energy*, 212, 997–1012.
- Athey, S., 2017. The impact of machine learning on economics. *Economics of Artificial Intelligence*. University of Chicago Press., <http://www.nber.org/chapters/c14009>.
- Bajpai, P. (2021). Top factors that affect the price of oil Available at <https://www.investopedia.com/articles/investing/072515/top-factors-reports-affect-priceoil.asp>, accessed January 15, 2021.
- Banna, H., Hassan, M. K., Ahmad, R., & Alam, M. R. (2022). Islamic banking stability amidst the COVID-19 pandemic: the role of digital financial inclusion. *International Journal of Islamic and Middle Eastern Finance and Management*, 15(2), 310-330.
- Chai, J., Xing, L.-M., Zhou, X.-Y., Zhang, Z.G., Li, J.-X., 2018. Forecasting the WTI crude oil price by a hybrid-refined method. *Energy Economics*, 71, 114–127.
- Collado, R., Creamer, G.G., 2016. Time series forecasting with a learning algorithm: an approximate dynamic programming approach. 22nd International Conference on Computational Statistics (COMPSTAT). pp. 111–122.

- Debnath, K.B., Mourshed, M., 2018. Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, 88, 297–325.
- D’Ecclesia, R. L., & Clementi, D. (2021). Volatility in the stock market: ANN versus parametric models. *Annals of Operations Research*, 299(1), 1101-1127
- Dogah, K.E., Premaratne, G., 2018. Sectoral exposure of financial markets to oil risk factors in BRICS countries. *Energy Economics*, 76, 228–256.
- Dudek, G., 2016. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32 (3), 1057–1060.
- Ftiti, Z., Tissaoui, K., & Boubaker, S. (2020). On the relationship between oil and gas markets: a new forecasting framework based on a machine learning approach. *Annals of Operations Research*, 1-29.
- Godarzi, A.A., Amiri, R.M., Talaei, A., Jamasb, T., 2014. Predicting oil price movements: a dynamic Artificial Neural Network approach. *Energy Policy* 68, 371–382.
- Gu, S., Kelly, B.T., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning, *Review of Financial Studies*, 33, (5), Pages 2223–2273.
- Hallack, M. and Weiss, M (2020). Electricity prices: the heterogeneous impact of COVID-19 on LAC markets. Available at <https://blogs.iadb.org/energia/en/electricity-prices-heterogeneousimpact-of-covid-19-in-lac-markets/>, accessed November 9, 2021
- Hallack, M., & Weiss, M. (2020). Electricity prices: The heterogeneous impact of COVID-19 on LAC markets Available at <https://blogs.iadb.org/energia/en/electricity-prices-heterogeneousimpact-of-covid-19-in-lac-markets/>, accessed January 16, 2021.
- Heckman, R. (2017). 10 factors that affect the cost of energy Available at <https://www.appenergy.com/2017/01/23/10-factors-that-affect-the-cost-of-energy/>, accessed January 15, 2021.
- Jia, C.-l., Wu, H.-s., Gong, D.-w., 2007. Coal demand prediction based on a support vector machine mode. *Journal of China University of Mining and Technology*, Ed. 36 (1), 107.
- Khosravi, A., Nahavandi, S., Creighton, D., 2013. Quantifying uncertainties of neural networkbased electricity price forecasts. *Applied Energy*, 112, 120–129.
- Li, J., Xu, Z., Xu, H., Tang, L., Yu, L., 2017. Forecasting oil price trends with sentiment of online news articles. *Asia-Pacific Journal of Operational Research*, 34 (02), 1740019.
- Linas, B., Savinkina, A., Barbosa, C., Mueller, P., Cerdá, M., Keyes, K. and Chhatwal, J., (2021). A clash of epidemics: Impact of the COVID-19 pandemic response on opioid overdose, *Journal of Substance Abuse Treatment*, 120, 108158, <https://doi.org/10.1016/j.jsat.2020.108158>.
- Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. *JOURNAL OF ECONOMIC PERSPECTIVES*, 31 (2), 87–106.

- Nelwamondo, F.V., Mohamed, S., Marwala, T., 2007. Missing data: a comparison of neural network and expectation maximization techniques. *Current Science*, 93 (11), 1514–1521.
- Norouzi, N., Zarazua de Rubens, G. Z., Enevoldsen, P., & Behzadi Forough, A. (2020). The impact of COVID-19 on the electricity sector in Spain: An econometric approach based on prices. *International Journal of Energy Research*, 1–13.
- Olanrewaju, B. T., Olubusoye, O. E., Adenikinju, A. F., & Akintande, O. J. (2019). A panel data analysis of renewable energy consumption in Africa. *Renewable Energy*, 140, 668–679.
- Olubusoye O., Akintande O., Yaya O., Ogbonna A., Adenikinju A., (2021) Energy pricing during the COVID-19 pandemic: Predictive information-based uncertainty indices with machine learning algorithm, *Intelligent Systems with Applications*, 12, 200050, <https://doi.org/10.1016/j.iswa.2021.200050>.
- Olubusoye, O. E., Ogbonna, A. E., Yaya, O. S., & Umolo, D. (2021). An Information-Based Index of Uncertainty and the predictability of Energy Prices. *International Journal of Energy Research*. <https://doi.org/10.1002/er.6512>.
- Panapakidis, I.P., Dagoumas, A.S., 2016. Day-ahead electricity price forecasting via the application of artificial neural network based models. *Applied Energy*, 172, 132–151.
- Papadimitriou, T., Gogas, P., Stathakis, E., 2014. Forecasting energy markets using support vector machines. *Energy Economics*, 44, 135–142.
- Raju, M. P., & Laxmi, A. J. (2020). IOT based online load forecasting using machine learning algorithms. *Procedia Computer Science*, 171, 551-560.
- Rasmussen, C. E. and C. K. I. Williams. 2006, *Gaussian Processes for Machine Learning*. MIT Press. Cambridge, Massachusetts.
- Salisu, A. A., & Akanni, L. O. (2020). Constructing a Global Fear Index for the COVID-19 Pandemic. *Emerging Markets Finance and Trade*. <https://doi.org/10.1080/1540496X.2020.1785424>.
- Samitas, A., Kampouris, I., Kenourgios, D., 2020. Machine learning as an early warning system to predict financial crisis. *International Review of Financial Analysis*, 71, 101507, <https://doi.org/10.1016/j.irfa.2020.101507>.
- Shamsudheen, S. V., Khattak, M. A., Muneeza, A., & Huda, M. (2022). COVID-19 and GCC stock market performance: an analysis of the boon (financial stimulus package) and curse (oil price plunge) effects. *International Journal of Islamic and Middle Eastern Finance and Management*, 15(2), 223-225.
- Singh, N., Mohanty, S.R., Shukla, R.D., 2017. Short term electricity price forecast based on environmentally adapted generalized neuron. *Energy*, 125, 127–139.
- Tang, L., Dai, W., Yu, L., Wang, S., 2015a. A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. *International Journal of Information Technology and Decision Making*, 14 (01), 141–169.

- The World Bank. (2020). Impact of COVID-19 on commodity markets heaviest on Energy Prices; lower oil demand likely to persist beyond 2021 Available at <https://www.worldbank.org/en/news/press-release/2020/10/22/impact-of-covid-19-oncommodity-markets-heaviest-on-energy-prices-lower-oil-demand-likely-to-persist-beyond2021>, accessed January 16, 2021.
- Tien, H. T., & Hung, N. T. (2022). Volatility spillover effects between oil and GCC stock markets: a wavelet-based asymmetric dynamic conditional correlation approach. *International Journal of Islamic and Middle Eastern Finance and Management*. Doi: 10.1108/IMEFM-07-20200370.
- Vapnik, V. 1995, *The Nature of Statistical Learning Theory*. Springer, New York.
- Wang, D., Luo, H., Grunder, O., Lin, Y., Guo, H., 2017. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Applied Energy*, 190, 390–407.
- Yang, J.H., Wu, C.Y., Zhang, G.M., 2014. Coal demand prediction in Shandong Province based on artificial firefly wavelet neural network. *Advanced Materials Research*. 962. pp. 1931–1935. Trans Tech Publ.
- Yazici, I., Beyca, O. F., Gurcan, O. F., Zaim, H., Delen, D., & Zaim, S. (2020). A comparative analysis of machine learning techniques and fuzzy analytic hierarchy process to determine the tacit knowledge criteria. *Annals of Operations Research*, 1-24.
- Yun-cai, N., 2003. A superposition wavelet-neural network model of coal demand forecast [J]. *Journal of China Coal Society*, 1, 021.
- Zhang, J.-L., Zhang, Y.-J., Zhang, L., 2015. A novel hybrid method for crude oil price forecasting. *Energy Economics*, 49, 649–659.
- Zhao, Y., Li, J., Yu, L., 2017. A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, 66, 9–16.
- Zhu, B., Shi, X., Chevallier, J., Wang, P., Wei, Y.-M., 2016. An adaptive multiscale ensemble learning paradigm for nonstationary and non-linear energy price time series forecasting. *Journal of Forecasting*, 35 (7), 633–651.

Appendix 1: Variable's description and sources

Variable name	Description	Source
Energy Equity Indices		
US Renewable Energy Equity (USRenewEnergy) Index	United States Renewable Energy Equity Index	Thomson Reuters DataStream
US Oil, Gas, and Coal Equity (USOilGasCoal) Index	United States Oil, Gas, and Coal Equity Index	Thomson Reuters DataStream
US Pipelines Equity (USPipelines) Index	United States Pipelines Equity Index	Thomson Reuters DataStream
US Oil Equity & Services (USOilEqSvs) Index	United States Oil Equity & Services Index	Thomson Reuters DataStream
US Oil & Gas Refining & Marketing Equity (USOilRefMkting) Index	United States Oil & Gas Refining & Marketing Equity Index	Thomson Reuters DataStream
US Oil: Crude Production Equity (USOilCrudeProd) Index	United States Oil: Crude Production Equity Index	Thomson Reuters DataStream
US International Oil & Gas (USIntOilGas) Index	United States International Oil & Gas Index	
US Energy (USEnergy) Index	Overall United States Energy Index	Thomson Reuters DataStream
Uncertainty Indices		

<p>Twitter Economic Uncertainty (TEU-USA) Index</p>	<p>The Twitter-based Economic Uncertainty Index exists since June 2011. Developed by Thomas Renault (University of Pantheon Sorbon, Paris 1) in collaboration with Scott R. Baker (Northwest), Nicholas Bloom (Stanford), and Steve Davis (University of Chicago). The method is as follows. Extracts all messages (tweets) sent on Twitter since June 2011 that contain keywords related to uncertainty and keywords related to business. The terms of uncertainty are: "'economists', 'economic', 'economy', economics', 'economically', 'economies', 'economist', 'economical.'"</p>	<p>www.policyuncertainty.com</p>
<p>EMV Infections Uncertainty (EMV) Index</p>	<p>EMV Infections Uncertainty is a daily newspaper-based infectious disease stock market volatility tracker. The index is daily and available from January 1985. To build an infectious disease EMV tracker, the author</p>	<p>www.policyuncertainty.com</p>
	<p>proceeds in four sentences: E: {financial, economy, economics}, "V: {uncertainty, risk, volatility, volatile}, M: {'Standard and Poors', 'equities', 'equity', 'stock market'}, "ID: {H1N1, epidemic, EBOLA, pandemic, MERS, virus, flu, SARS disease, coronavirus, H5N1}. Then, in about 3,000 US newspapers, get the number of daily newspaper articles that contain at least one term in E, M, V, and ID. Third, they scale the raw EMVID count by the number of items on the same day. The final step contains the rescaling of the resulting series. This reflects the approach of scaling the category's EMC series with the category's EMC tracker.</p>	
<p>Economic Policy Uncertainty (EPU) index</p>	<p>The Daily news-based Economic Policy Uncertainty Index includes newspaper archives from Access World News's NewsBank service. The NewsBank Access World News database is based on an archive of newspapers and other news sources around the world. NewsBank has a wide range of news sources, from newspapers to magazines to news agencies, but the analysis is done using only newspaper sources. The index is limited to US newspapers, and NewsBank has well over 1,000 newspapers. These newspapers range from large national newspapers such as USA Today to small local newspapers nationwide.</p>	<p>www.policyuncertainty.com</p>

CBOE Market Volatility Index (VIX)	Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX) VIX is the ticker symbol and the popular name for the Chicago Board Options Exchange's CBOE Volatility Index, a popular measure of the stock market's expectation of volatility based on S&P 500 index options.	Thomson Reuters DataStream
------------------------------------	---	----------------------------