Journal of Climate Finance 1 (2022) 100002

Contents lists available at ScienceDirect

Journal of Climate Finance

journal homepage: www.journals.elsevier.com/journal-of-climate-finance

Using machine learning to predict clean energy stock prices: How important are market volatility and economic policy uncertainty?

Perry Sadorsky

Schulich School of Business, York University, 4700 Keele Street, Toronto M3J 1P3, Ontario, Canada

ARTICLE INFO

JEL Classification:
G17
Q42, Q47

Keywords:
Clean energy stock prices
Forecasting
Machine learning
Random forests
VIX

ABSTRACT

The disruptive impacts of climate change have created an urgent need to transition to a low carbon economy and an important part of this transition is an increase in the usage of clean energy. The greater adoption of clean energy is creating new opportunities for clean energy equity investing. The existing literature mostly focuses on the dynamic relationship between clean energy equities, oil prices, technology stock prices, and other important macroeconomic variables like market volatility and economic policy uncertainty. However, there is a shortage of literature on forecasting clean energy stock prices. Forecasting clean energy equity prices is important for making investment decisions. This paper uses machine learning methods to predict the direction of clean energy stock prices. The analysis reveals that random forests, extremely randomized trees, stochastic gradient boosting, and support vector machine have higher prediction accuracy than Lasso or Naïve Bayes. For forecasts in the 10-day to 20-day range, random forests, extremely randomized trees, stochastic gradient boosting, and support vector machine achieve prediction accuracies greater than 85 %. In some cases, prediction accuracy reaches 90%. Lasso prediction accuracy is higher than Naïve Bayes but never greater than 65 %. The MA200, MA50, and WAD technical indicators are, on average, the features most important for predicting clean energy stock price direction. Of the non-technical indicators, VIX and OVX are consistently ranked high in importance. In most cases, EPU is not one of the most important features, Of the forecasting methods considered, extremely randomized trees are very impressive due to high accuracy and short computational time.

1. Introduction

The disruptive impacts of climate change have created an urgent need to transition to a low carbon economy. An important part of this transition to a low carbon economy is to increase the usage of clean energy relative to fossil fuels. According to BloombergNEF (BloombergNEF, 2022) global energy transition investment in 2021 amounted to $755 billion (US). This investment includes renewable energy, carbon capture and storage (CCS), nuclear, hydrogen, sustainable materials, electrified heat, electrified transport, and electricity storage. Over the period 2004–2021 the compound annual growth rate (CAGR) of energy transition investment was 20%. For the year 2021 the two largest sectors were renewable energy and electrified transport. Renewable energy growth was due to large increases in wind and solar installations while the growth in electrified transport was due to large increases in electric vehicle sales. At current trends, the electrified transport sector will overtake renewable energy as the top sector by the year 2024. In 2021, investment levels increased to a new record in every major region of the world. On a country basis, China's energy transition investment in 2021 was the highest ($268 billion), the US was second ($114 billion), and Germany was third ($47 billion).

Increases in clean energy usage are creating new opportunities for clean energy equity investing. The existing clean energy finance literature mostly focuses on the dynamic relationship between the stock prices of clean energy companies, oil prices, the stock prices of technology companies, and other important macroeconomic variables like financial market volatility and economic policy uncertainty (Bondia et al., 2016; Dutta, 2017; Dutta et al., 2018; Elie et al., 2019a; Elie et al., 2019b; Ferrer et al., 2018; Geng et al., 2021; Gupta, 2017; Henriques & Sadorsky, 2008; Kumar et al., 2012; Le et al., 2021; Maghyereh et al., 2019; Managi & Okimoto, 2013; Naeem et al., 2020; Nasreen et al., 2020; Pham, 2021, 2019; Reboredo, 2015; Reboredo et al., 2017a, b; Reboredo & Ugolini, 2018; Saeed et al., 2021; Uddin et al., 2019; Wen et al., 2014). These studies show that oil prices and technology stock prices affect clean energy stock prices while the impact of economic policy uncertainty on the stock prices of clean energy companies is mixed. However, forecasting the prices of clean energy stocks is a topic that is understudied. Forecasting clean energy equity prices is imperative for making well informed investment decisions about this growing asset class. This raises several important questions. Which forecasting methods are the best to use? How important are market

E-mail address: psadorsky@schulich.yorku.ca.

https://doi.org/10.1016/j.jclimf.2022.100002
Received 21 October 2022; Received in revised form 16 November 2022; Accepted 21 November 2022

volatility and economic policy uncertainty (EPU) for predicting clean energy stock prices? Is variable importance consistent across different forecast horizons? These are the questions addressed in this paper.

In answering these questions, the following approach is followed. First, this paper focusses on predicting the direction of clean energy stock prices. Predicting the direction of clean energy stock prices is relevant because in practice, investors are more interested in the direction of asset prices rather than the actual values when determining asset allocation (Pesaran & Timmermann, 2002). Numerous studies have reported high accuracy from predicting asset price direction (Ballings et al., 2015; Basak et al., 2019; Leung et al., 2000; Lohrmann & Luukka, 2019; Nyberg, 2011; Nyberg & Pönkä, 2016; Pönkä, 2016; Sadorsky, 2022). Second, machine learning methods are used to predict the price direction of clean energy stocks. Compared to standard regression methods, machine learning methods can achieve higher prediction accuracy when the relationship between the predicted variable and features is complex (Hastie et al., 2009; James et al., 2013; Mullainathan & Spiess, 2017). Building on this existing literature, the machine learning techniques used in this present paper include random forests, extremely randomized trees, stochastic gradient boosting, support vector machine, Lasso, and Naïve Bayes. Third, the features (also referred to as explanatory variables or predictors) includes technical indicators, market volatility and policy uncertainty. The relevance of technical indicators for asset price prediction has been well established (Bustos & Pomares-Quimbaya, 2020; Neely et al., 2014; Sadorsky, 2022, 2021a,b; Wang et al., 2020; Yin et al., 2017; Yin & Yang, 2016). Including market volatility and policy uncertainty is a new contribution to the literature on predicting clean energy stock prices.

Several important findings emerge from the analysis reported in this paper. First, random forests, extremely randomized trees, stochastic gradient boosting, and support vector machines have much higher forecast accuracy than either Lasso or Naïve Bayes when used to predict the direction of clean energy equity prices. These results add support to the existing literature demonstrating the high prediction accuracy of machine learning models for predicting stock price direction. Second, the analysis from this paper finds that the technical indicators MA200, WAD, and MA20 are important features for predicting the prices of clean energy stocks. This result is consistent with the literature showing the importance of technical indicators for forecasting stock prices. The stock market volatility (VIX) and oil price volatility (OVX) are important non-technical predictors. Third, for 15 day forecasts, the most important features are, in most cases, also the most important features for 20 day forecasts. This is relevant since the highest prediction accuracy is observed in the 15–20 day range.

This paper is organized as follows. The next section of the paper presents the background literature. Section 3 describes the machine learning methods used in the analysis. The data are described in Section 4. The results are reported in Section 5. Section 6 concludes the paper and offers some practical implications stemming from the results.

2. Background literature

Stock market volatility can have a major impact on stock prices because higher volatility implies more risk (Schwert, 1990). For example, investments that looked attractive in a low risk environment may not be as attractive in a high risk environment. Oil price volatility and technology stock market volatility may also have an important impact on clean energy stock prices. Since clean energy and fossil fuels are substitutes, higher oil prices should encourage a shift away from fossil fuels to clean energy. Higher oil price volatility can, however, create delays in decision making because of uncertainty over the future profitability of renewable energy investments. Clean energy relies on technological innovation to bring down the cost of generating energy from renewable sources and uncertainty about technology can delay the decision to invest in clean energy. This uncertainty can be captured in technology stock market volatility (Henriques & Sadorsky, 2008).

Dutta (2017) discovers a positive relationship between the realized volatility of clean energy equities and oil price volatility (OVX). Ahmad et al. (2018) find a negative correlation between renewable energy stock volatility and OVX. In studying the stock price dynamics of solar energy firms, Dutta (2019) finds that OVX has a larger impact on solar stock prices than either gold price volatility or silver price volatility. Dutta (2018) finds that OVX is useful for hedging the downside risk of clean energy stock prices. Sadorsky (2022) finds that the volatility of oil prices along with the volatility of silver prices are important predictors of solar energy stock prices. Dutta et al. (2020) study how useful gold, silver, and oil volatility indices are for hedging downside risk of clean energy equities. They find that the crude oil volatility index is the most effective hedging asset followed by gold and silver.

Economic policy uncertainty in the form of uncertainty about fiscal policy, monetary policy and regulatory uncertainty can affect business decisions and financial markets (Baker et al., 2016). In the face of uncertainty, businesses are reluctant to commit to capital spending which affects current and future productivity and profitability (Pástor & Veronesi, 2012). An increase in economic policy uncertainty can indicate worsening economic performance which will affect the cost of capital and other real option decisions (Arouri et al., 2016).

Lundgren et al. (2018) find that uncertainty variables (VIX and financial stress) have a significant impact on clean energy stock returns and volatility. During the global financial crisis and the European debt crisis uncertainty is mostly a net transmitter of connectedness. Ferrer et al. (2018) using time and frequency connectedness and controlling for the impact of financial factors find that clean energy stocks are similar to technology stocks. In comparing the impact of financial market uncertainty, oil market uncertainty, and economic policy uncertainty on clean energy stocks, Ji et al. (2018) find that economic policy uncertainty has the weakest effect. Liu and Hamori (2020) find that the VIX has a significant impact on renewable energy stocks and the impact of stock market uncertainty on clean energy stocks is greater in the US as compared to Europe. Zhao (2020) finds that economic policy uncertainty shocks have a negative impact on clean energy stock prices. Chakrabarti and Sen (2021) study the market risk of green energy equities. They find that green stocks are defensive and may be of interest to investors who want protection against downside risk. Volatility spills over from domestic markets to green equities. Market risks and spillovers are more variable in the US and Europe than they are in the Asia Pacific. Green stocks have shown resilience to the COVID-19 period. Saeed et al. (2021) find that economic policy uncertainty has no statistically significant impact on the total spillover index for clean energy stocks, green bonds, crude oil, and fossil fuel energy stocks. Liu et al. (2022) find that the impact of economic policy uncertainty on clean energy stocks in the US, Europe, and the world tends to be concentrated at a high frequency. Uncertainty caused by COVID-19 is more significant than that caused by the global financial crisis.

This present paper uses machine leaning techniques to predict clean energy stock prices because machine learning techniques like random forests and support vector machines have demonstrated high accuracy when predicting stock prices (Ampomah et al., 2020; Ballings et al., 2015; Basak et al., 2019; Ghosh et al., 2021; Khan et al., 2020; Lohrmann & Luukka, 2019; Weng et al., 2018). The research that comes closest to this present study are the papers by Sadorsky (2021a) and Sadorsky (2022). Sadorsky (2021a) uses random forests to predict the direction of clean energy stock prices. He finds that random forests and tree bagging have higher predictive accuracy than logit models. The feature space includes technical indicators but does not include business cycle variables. Sadorsky (2022) uses random forests and support vector machine to predict solar stock prices. Random forests, extremely randomized trees, tree bagging, and support vector machine produce more accurate predictions than logit or boosted logit. The feature space includes technical indicators, the volatility of oil and silver prices and silver prices. The most important predictors are technical indicators. Non-technical features like the volatility of oil prices and the volatility of silver prices rank in the top third. Neither of these two studies include a more comprehensive set of market volatility or policy uncertainty variables.

In summarizing the literature, it appears that oil price volatility has a larger impact on clean energy stock prices than does stock market volatility. The impact of EPU on clean energy stock prices is mixed. There is, however, little known about how useful market volatility and economic policy uncertainty are for forecasting clean energy stock prices. This is the gap in the literature that this paper addresses.

3. Methods

3.1. Machine learning methods

This section provides a brief non-technical description of the machine learning methods (random forests, boosting, extremely randomized trees, support vector machine, Naïve Bayes, and Lasso) used to predict clean energy stock price direction. James et al. (2013) provide a very informative and easy to read overview of these methods while a more technical description is provided by Hastie et al. (2009).

Random forests and extremely randomized trees are machine learning methods based on ensembles of decision trees. Decision trees can be trained to have high prediction accuracy on a training data set, but decision trees are susceptible to high variance. High variance is a problem because small changes in the data can produce different predictions. High variance can be addressed by using ensemble methods that create many de-correlated trees and then average the predictions across the trees. Random forests and extremely randomized trees (Extra Trees) are two examples of decision tree ensemble methods. Random forests are collections of decision trees created using bootstrapping (Breiman, 2001). Many training samples are created using bootstrapping and a decision tree is fit to each bootstrapped sample. A random sample of predictors is chosen from the full set of predictors each time a split in a tree occurs. The random choice of features helps to reduce the correlation between the trees (James et al., 2013). The predictions from each tree are averaged to provide an overall prediction. Extra Trees are similar to random forests in that a random set of features is chosen at each split but unlike random forests the split values for these features are selected at random (Geurts et al., 2006). The best split is chosen from this small number of choices. Extra Trees does not use bootstrapping but samples from the original data set. Extra Trees creates an ensemble of trees containing trees that have more variability but less correlation than the trees from random forests (Geurts et al., 2006). Compared to random forests, Extra Trees takes much less computational time because splits are chosen at random and bootstrapping is not used.

Instead of creating an ensemble of trees, tree boosting makes a sequence of small adjustments to one decision tree (James et al., 2013). A boosted decision tree is constructed as follows. First, a decision tree is fit to the data. Then a decision tree is fit to the residuals. This new decision tree is added to the fitted function. The updated residuals are again used as the dependent variable in a decision tree and the fitted function further updated. This sequential approach of updating the fitted function continues until a stopping criterion is met. The estimation process fits small trees to the residuals at each step and in this way slowly improves the fit of the original decision tree. There are three main tuning parameters for boosting; the number of trees, the shrinkage parameter which controls the rate at which learning takes place, and the number of splits in each tree (interaction depth) which controls the complexity of the tree. There is an inverse relationship between the learning rate and the number of trees and in many practical applications boosting can take much more computational time than random forests.

Support Vector Machine (SVM) is a machine learning method that partitions a data set into separate groups. The separation between groups is achieved using a boundary (hyperplane) (James et al., 2013). There are many different hyperplanes which can be used to separate the data and the goal of SVM is to find the maximum margin hyperplane (MMH). This is the hyperplane that has the greatest separation between the groups of data. The points that are closest to the MMH are the support vectors. For linearly separable data the support vectors can be found easily. For more complex problems the support vectors can be located by using advanced vector geometry. Kernels can be used to map the data into higher dimensionally space. Mapping the data into higher dimensional space creates the possibility that a nonlinear relationship in a specific dimension may become linear in higher dimensional space.

Two classification methods, Naïve Bayes and Lasso are included as benchmarks (James et al., 2013). Modeling stock price direction is a classification problem because a binary variable can be used to indicate the direction of stock price change. For the analysis reported in this paper, stock price change from one period to the next can be classified as either positive or non-positive. The Naïve Bayes classifier uses Bayes theorem to make predictions. The Naïve Bayes classifier is a simple probability classifier that is often used as a benchmark.

The least absolute shrinkage and selection operator (Lasso) is a regression based method that does regularization and variable selection. Lasso has similarities with ridge regression (James et al., 2013). Whereas ridge regression shrinks parameter estimates down to small but non-zero values, Lasso may set parameter values equal to zero which results in a sparse model. The shrinkage of coefficients reduces variance and bias.

3.2. Model specifics

For each model, multi-step forecasts are made from one day to twenty days. A multi-step forecast horizon provides additional information on how model accuracy changes with the length of forecasts. The choice of twenty days is based on the fact that twenty days is approximately the average number of trading days in a month. As is usual practice in machine learning applications, the data were randomly split into a training set consisting of 70 % of the data and a testing set consisting of 30 % of the data. For random forests and extremely randomized trees there are two main tuning parameters; the number of trees and the number of randomly selected features. At each split the number of predictors was randomly chosen using the formula the square root of the total number of features (James et al., 2013). The number of trees was set at five hundred. The number of trees does not affect the prediction accuracy of random forests so long as a large enough number of trees are used. A small number of trees, however, results in high test error. In the case of a 20-day random forests forecast for the clean energy ETF PBW, the test error decreases sharply as the number of trees closes in on 200 trees. The test error does not change much as the number of trees passes 200. This was also the case for other forecast horizons. Results for the ETFs ICLN and QCLN are similar to that of PBW. For the Extra Trees, one randomly chosen split point is used for each randomly selected feature. Notice the defining element of extremely randomized trees; the number of features is chosen randomly and the split for each feature is also chosen randomly.

The stochastic gradient boosted model used in this paper has several tuning parameters including the number of trees, the interaction depth, and shrinkage. The tuned model for 20 day forecasts of PBW was estimated with 1000 trees, a bag fraction of 0.5, an interaction depth of 4, a minimum number of observations in each node of 10, and a shrinkage of 0.1. This model also had high prediction accuracy for the ICLN and QCLN ETFs. These values were chosen based on bootstrap re-sampling of a grid search with the number of trees (500, 1000, 2000, 3000), the interaction depth (2, 4, 5) and shrinkage (0.01, 0.1, 0.2). Stochastic gradient boosting can take up a lot of computational time, so the model was not trained on the other forecast horizons.

The SVM was estimated with a radial basis function and two tuning parameters (cost and gamma). Ten-fold cross validation with 10 repeats was used to determine the optimal values for each forecast. The tuning grid for the cost parameter contained the values (0.1, 1, 10, 100, 1000). The tuning grid for the gamma parameter contained the values (0.2, 0.5, 1, 2, 3, 4). Distance-based algorithms like the ones used by SVM are susceptible to the range of the data. Consequently, the features were scaled to have a mean of zero and standard deviation of one. Data scaling was not applied to the tree-based methods because splitting occurs on a single feature which is not affected by scaling.

For classification problems the confusion matrix provides the main output from which to create measures of prediction accuracy. The number of true positives and true negatives divided by the total number of predictions determines the prediction accuracy. Prediction accuracy ranges between zero and one with one being a perfect fit. Two other prediction accuracy measures are included; kappa and FI. The kappa statistic is a better measure of prediction accuracy when there is a considerable unbalance between the classification categories because kappa adjusts prediction accuracy by accounting for the possibility of a chance occurrence of a correct prediction. The F1 score, sometimes called the F measure, is a prediction accuracy measure that combines precision and recall into a single number which also takes into account class imbalance.

Shapley values are used to explain feature importance. Shapley values are derived from conditional game theory. Essentially, the Shapley value signifies the contribution of each feature towards the predicted value compared to the average prediction for the data set (Lundberg & Lee, 2016, 2017). As an example, consider the Shapley calculation for one feature say X1. The accuracy of every combination of features not

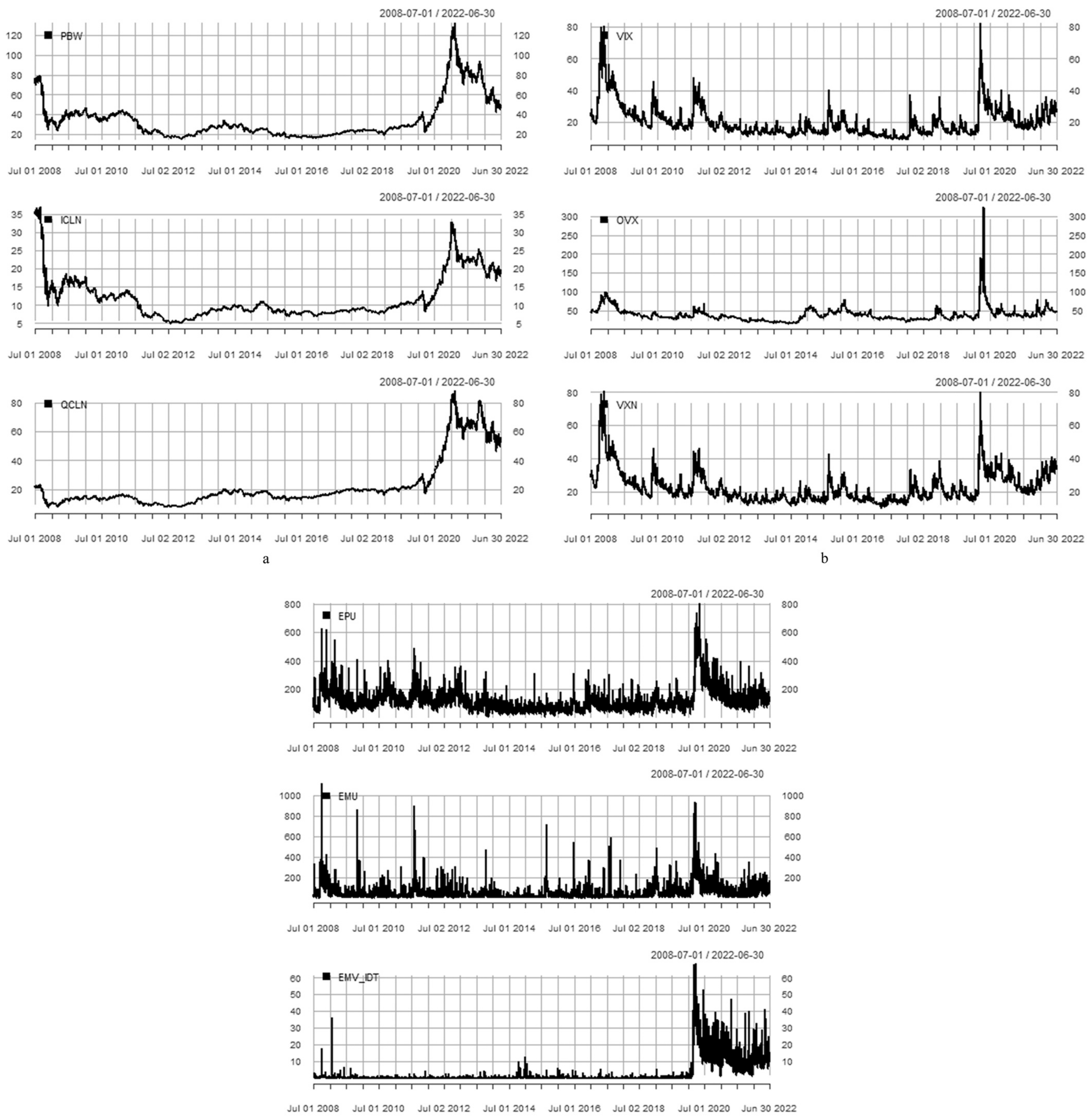

Fig. 1. (a). The time series of clean energy ETFs.b. The time series of financial market volatility.c. The time series of uncertainty indicators.

Table 1
Summary statistics.

	median	mean	std.dev	coef.var	skewness	kurtosis	normtes	normtest.p
PBW	0.085	-0.015	2.351	-161.482	-0.411	5.278	0.936	0.000
ICLN	0.000	-0.018	2.159	-120.305	-0.619	10.580	0.878	0.000
QCLN	0.135	0.024	2.218	92.522	-0.432	5.431	0.937	0.000
VIX	17.355	20.234	9.747	0.482	2.333	7.431	0.783	0.000
OVX	35.045	39.014	19.090	0.489	4.253	34.561	0.705	0.000
VXN	19.700	22.603	9.433	0.417	2.060	6.073	0.818	0.000
EPU	101.570	124.721	86.497	0.694	2.333	8.210	0.802	0.000
EMU	32.750	59.908	82.233	1.373	4.150	28.319	0.615	0.000
EMV_IDT	0.350	3.108	7.363	2.369	3.536	15.658	0.483	0.000

PBW, ICLN, and QCLN are expressed in continuous compound returns. Other variables are in original units. Data are for the period July 2, 2008 to June 30, 2022.

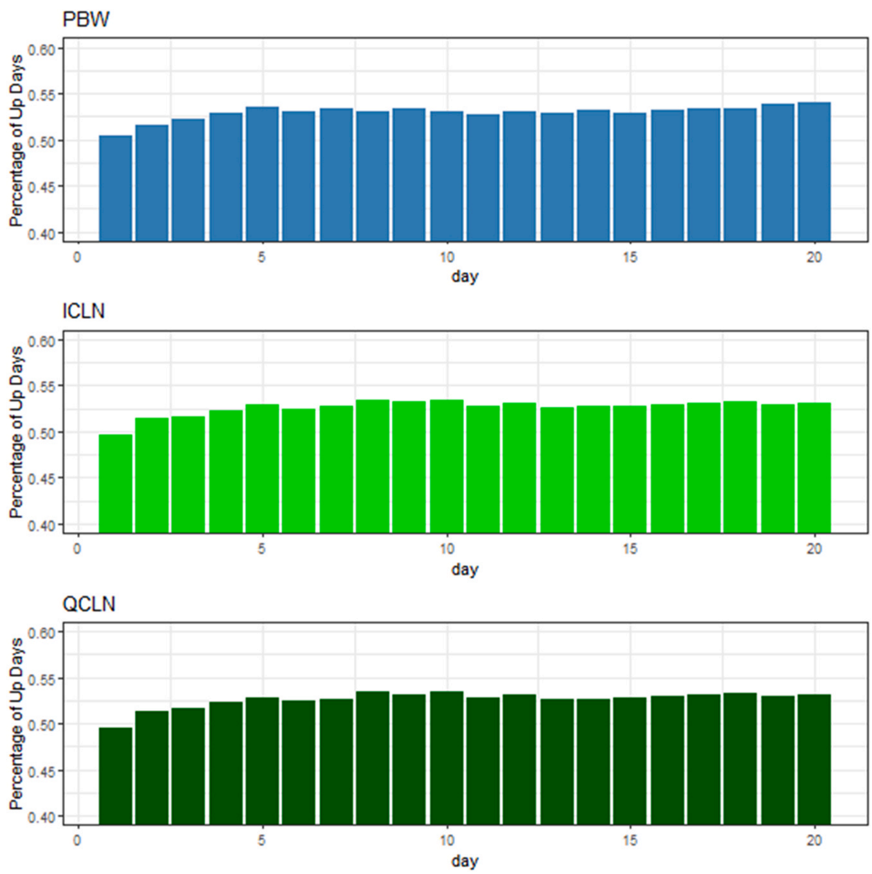

Fig. 2. The percentage of positive days.

including X1 is recorded and then a test is conducted to see how adding X1 to each combination improves the accuracy. Shapley values are theoretically consistent and more robust than permutation based variable importance. The tradeoff is that Shapley values are much more computationally intensive and have a much longer computation time.

All calculations were done in R (R Core Team, 2022) using the ranger machine learning package (Wright et al., 2022), the e1071 package (Meyer et al., 2021), the Extra Trees package (Simm & Abril, 2014), the GBM package (Greenwell et al., 2020), the caret package (Kuhn et al., 2020) and the fastshap package (Greenwell, 2021).

4. Data

The study used data on clean energy stock prices, stock market volatility, oil market volatility, technology stock market volatility, economic market uncertainty, economic policy uncertainty, and the equity market infectious disease tracker. Clean energy stock prices are measured using three of the most widely traded clean energy exchange traded funds (ETFs). The Invesco WilderHill Clean Energy ETF (PBW) is based on the WilderHill Clean Energy Index. The WilderHill Clean Energy Index is a collection of stocks of publicly traded companies listed on a major exchange in the U.S. that are involved in advancing cleaner energy and energy conservation. PBW holds about 80 stocks. As of April 30, 2022, the top 10 holdings comprised 22% of total assets. The second clean energy ETF is the ICLN. The ICLN is based on the iShares Global Clean Energy Index and tracks the performance of approximately 100 global clean energy-related companies. As of April 30, 2022, the top 10 holdings accounted for 49 % of total assets. The third clean energy ETF is QCLN which is based on the NASDAQ Clean Edge Green Energy Index. This index consists of small, mid and large capitalization clean energy companies that are publicly traded in the United States. QCLN typically holds about 65 stocks. As of April 30, 2022, the top 10 holdings consisted of 56 % of total assets.

The CBOE VIX index, which tracks the S&P 500 volatility, measures US stock market volatility. The VIX is calculated using a complicated formula that averages the weighted prices of out of the money puts and

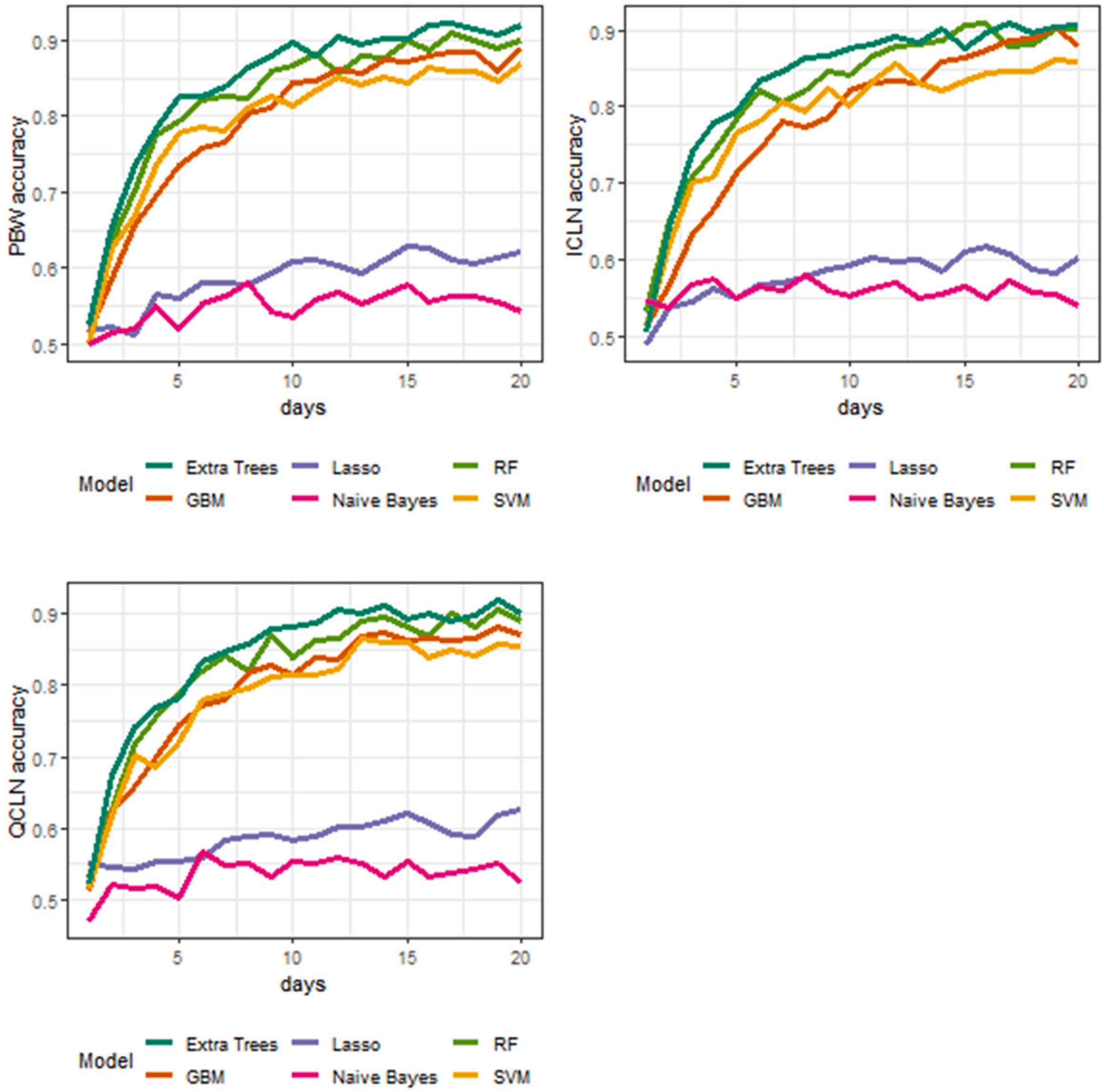

Fig. 3. Prediction accuracy.

calls that expire in 16 and 44 days on the S&P 500. Technology stock market volatility is represented by VXN which is computed in a similar way as VIX but for the NASDAQ 100 stocks. The CBOE oil price volatility index (OVX) is used to measure oil price volatility. Options prices on crude oil futures are used to create the OVX.

US equity market uncertainty (EMU), US economic policy uncertainty (EPU) and the infectious disease equity market volatility (EMV_IDT) are obtained from the economic policy uncertainty website (https://www.policyuncertainty.com/) (Baker et al., 2016). These indices are based on key word searches in newspapers and media.

The data for this study is daily and starts on July 1, 2008 (ICLN began trading on June 24, 2008) and ends on June 30, 2022. Stock price and market volatility data were collected from Yahoo Finance. Data on the uncertainty variables was obtained from FRED (https://fred.stlouisfed.org/). The stock price data are used to construct technical indicators. The technical indicators used in this study include two moving averages (the 50-day and 200-day), the relative strength indicator (RSI), the Williams accumulation and distribution (WAD) indicator, stochastic oscillators (slow (StoSLOWD), fast(StoFASTK, StoFASTD)), advance – decline line (ADX), moving average cross-over divergence (MACD), price rate of change (ROC), on balance volume (OnBalanceVolume), and the money flow index (MFI). These widely used technical indicators have demonstrated predictive power for forecasting stock prices (Bustos & Pomares-Quimbaya, 2020; Neely et al., 2014; Wang et al., 2020; Yin et al., 2017; Yin & Yang, 2016). The default settings in the R package TTR (Ulrich, 2020) were used in calculating the technical indicators.

Time series patterns of the clean energy ETFs are shown in Fig. 1a. Clean energy stocks, like most stocks, dropped considerably during the 2008–2009 global financial crisis. Except for 2014, clean energy stock prices did not fluctuate very much between January 2, 2013 and January 2, 2019. The year 2014 was important because global investment in clean energy increased 16% that year mostly due to increased solar power in the US and China. Wind power investment in Europe was also strong.[1] After January 2019, clean energy stock prices increased and peaked in early January of 2021. This time period reflects the onset of the COVID19 pandemic. After January of 2021 prices declined somewhat but as of March 2022 are still above the values recorded in 2019.

[1] https://www.theguardian.com/environment/2015/jan/09/solar-power-drives-renewable-energy-investment-boom-2014

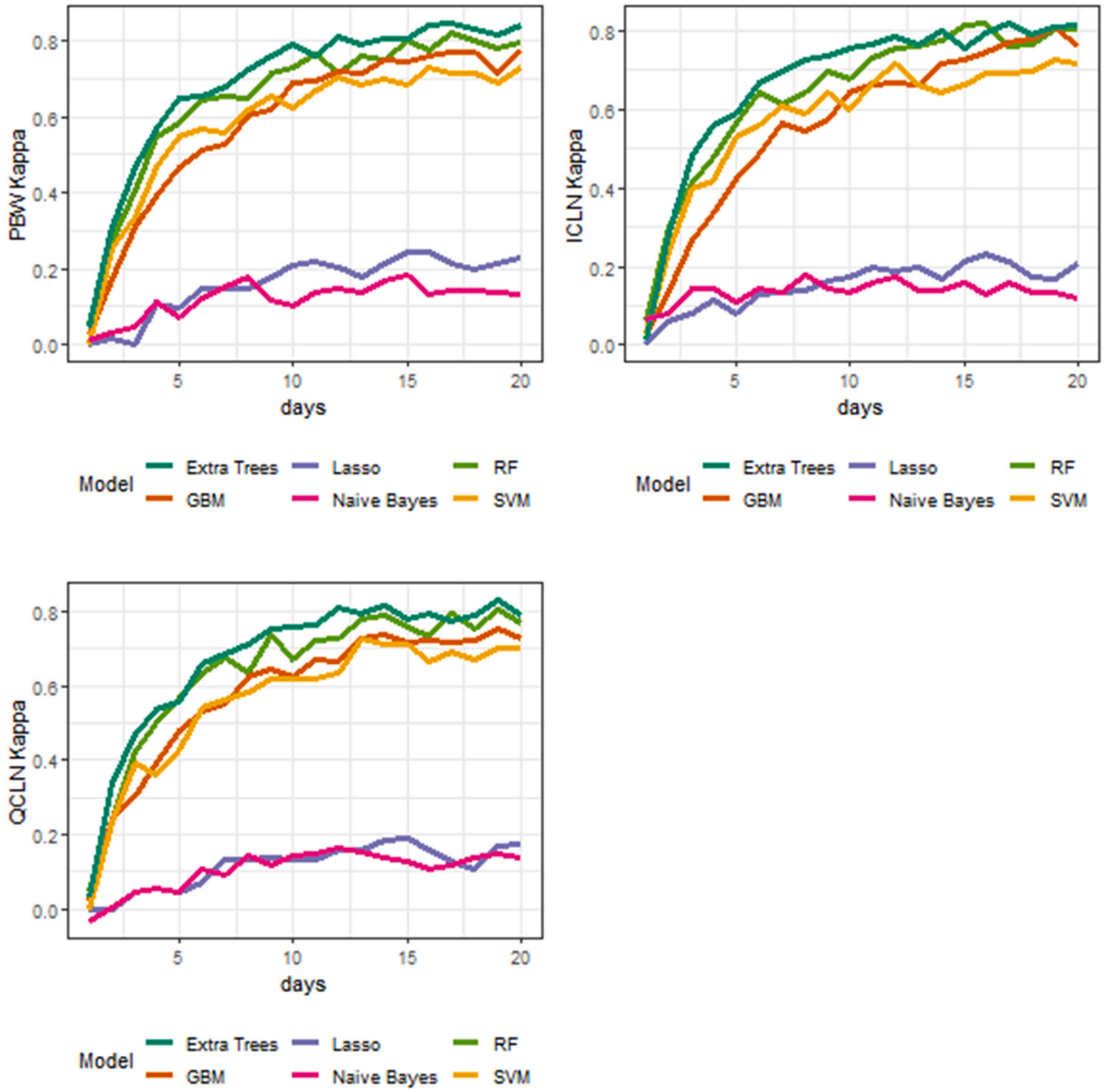

Fig. 4. Kappa values.

VIX and VXN display similar time paths (Fig. 1b). These market volatility indices recorded high values during the 2008–2009 global financial crisis and the beginning of the COVID19 pandemic (March 2020). Notice that in March of 2020, there was a large increase in oil price volatility. This was when the price of the front month oil futures contract briefly turned negative.

Equity market uncertainty and economic policy uncertainty recorded high values during the global financial crisis and March 2020 (beginning of the COVID19 pandemic) (Fig. 1c). The COVID19 pandemic had been detected in late 2019 and was a topic of concern in early 2020 but it was declared a global pandemic by the World Health Organization in March 2020. This is reflected in the time pattern of the equity market volatility infectious disease tracker (EMV_IDT).

Of the three ETFs, only QCLN had a positive average return over the sample period (Table 1). The coefficient of variation values indicates that OVX is more variable than VIX and VIX is more variable than VXN. Among the uncertainty indices, EMV_IDT has the most variation and EPU is the least variable. The data display non-normality as for each variable the median is different from the mean, kurtosis and skewness are present, and the null hypothesis of normality is rejected.

The percentage of up (positive) days for each forecast day and each clean energy ETF are displayed in Fig. 2. The values range between 0.50 and 0.55. This indicates a relatively even class distribution.

5. Results

This section of the paper reports the results from the analysis. Looking first at the accuracy for predicting PBW, random forests, stochastic gradient boosting, extremely randomized trees and SVM produce higher prediction accuracy than either Lasso or the Naïve Bayes models (Fig. 3). The prediction accuracy of random forests (RF),

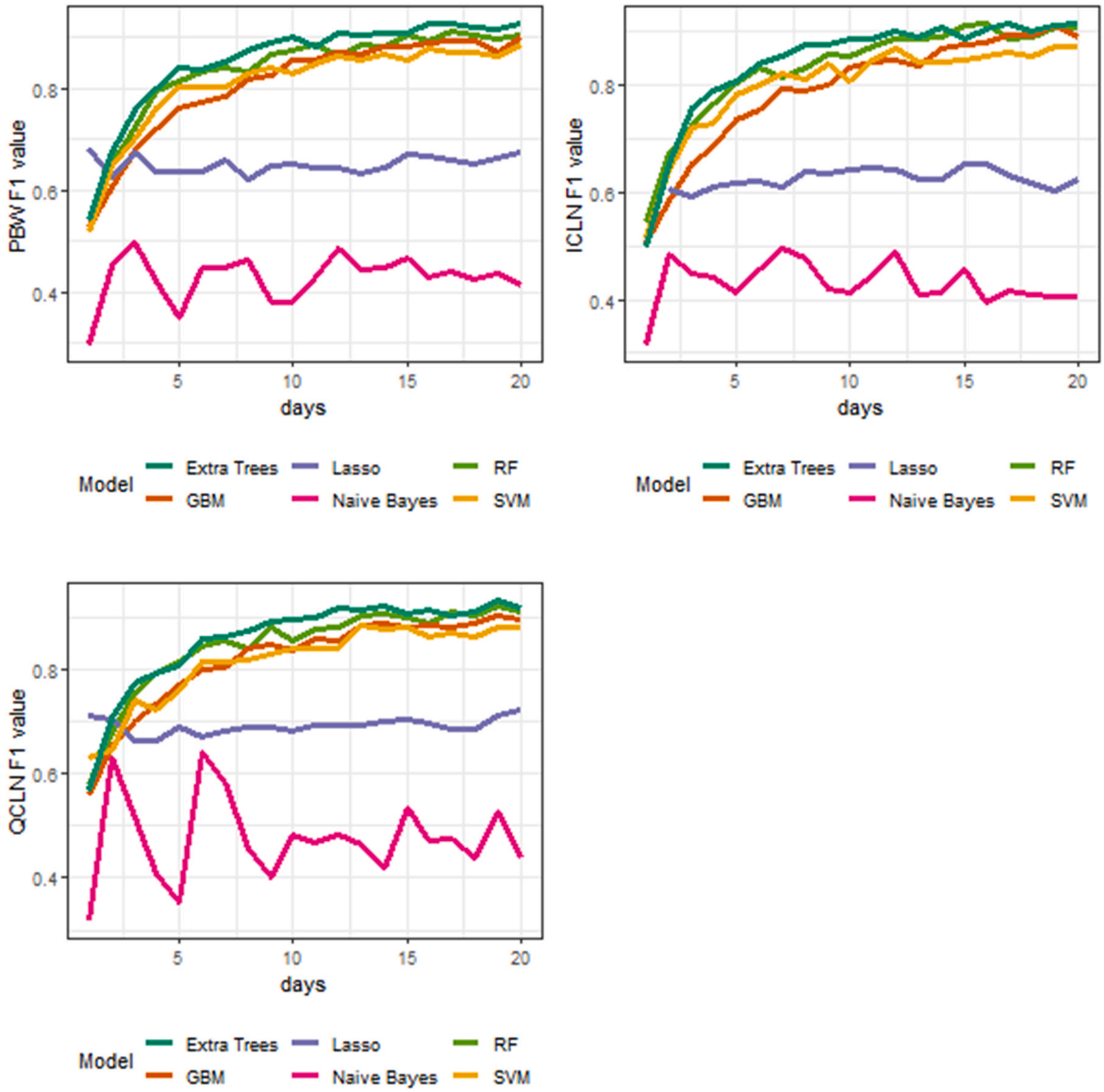

Fig. 5. F1 value.

stochastic gradient boosting model (GBM), SVM, and extremely randomized trees (Extra Trees) increases quickly up to eight days. After 10 days the accuracy of these methods is over 80 %. Between 15 and 20 days, random forests, GBM, and extremely randomized trees obtain accuracies of 90 % or in some cases slightly greater. The pattern of accuracy observed for PBW across the number of days forecast is similar to that of ICLN and QCLN. Overall, random forests, GBM, and Extra Trees have the highest accuracy followed by SVM, Lasso, and Naïve Bayes. These results are supportive of the findings by Sadorsky (2021a) who finds that random forests have high accuracy for predicting the price direction of clean energy ETFs.

Prediction accuracy measured using kappa (Fig. 4) shows that random forests, GBM, SVM, and Extra Trees have the highest kappa values. In comparison, Lasso and Naïve Bayes have the lowest kappa values. The pattern of kappa values reported in Fig. 4 is like that of the accuracy measures reported in Fig. 3.

After 5-days, the F1 values for random forests, GBM, Extra Trees and SVM are mostly greater than 0.80 (Fig. 5). By comparison, Naïve Bayes has the lowest F1 values and Lasso has the second lowest F1 values. In summary, overall accuracy, kappa, and F1 are in agreement showing random forests, GBM, Extra Trees and SVM have higher prediction accuracy than Lasso or Naïve Bayes.

Turning next to the importance of the features, mean absolute Shapley values for PBW are reported in Fig. 6a. Fig. 6a shows four subplots each representing a different forecast period. For each plot in Fig. 6a, features are ordered in descending order of importance. The horizontal axis denotes the probability associated with the up classification. For a 10-day forecast, for example (Fig. 6a, top right), WAD increases the probability of being in the up classification by 3.25 % on average. For a twenty day forecast, WAD increases the probability of being in the up classification by 8 % on average (Fig. 6a, bottom right). For forecasts of 10, 15, and 20 days, WAD, MA50, and MA200 are

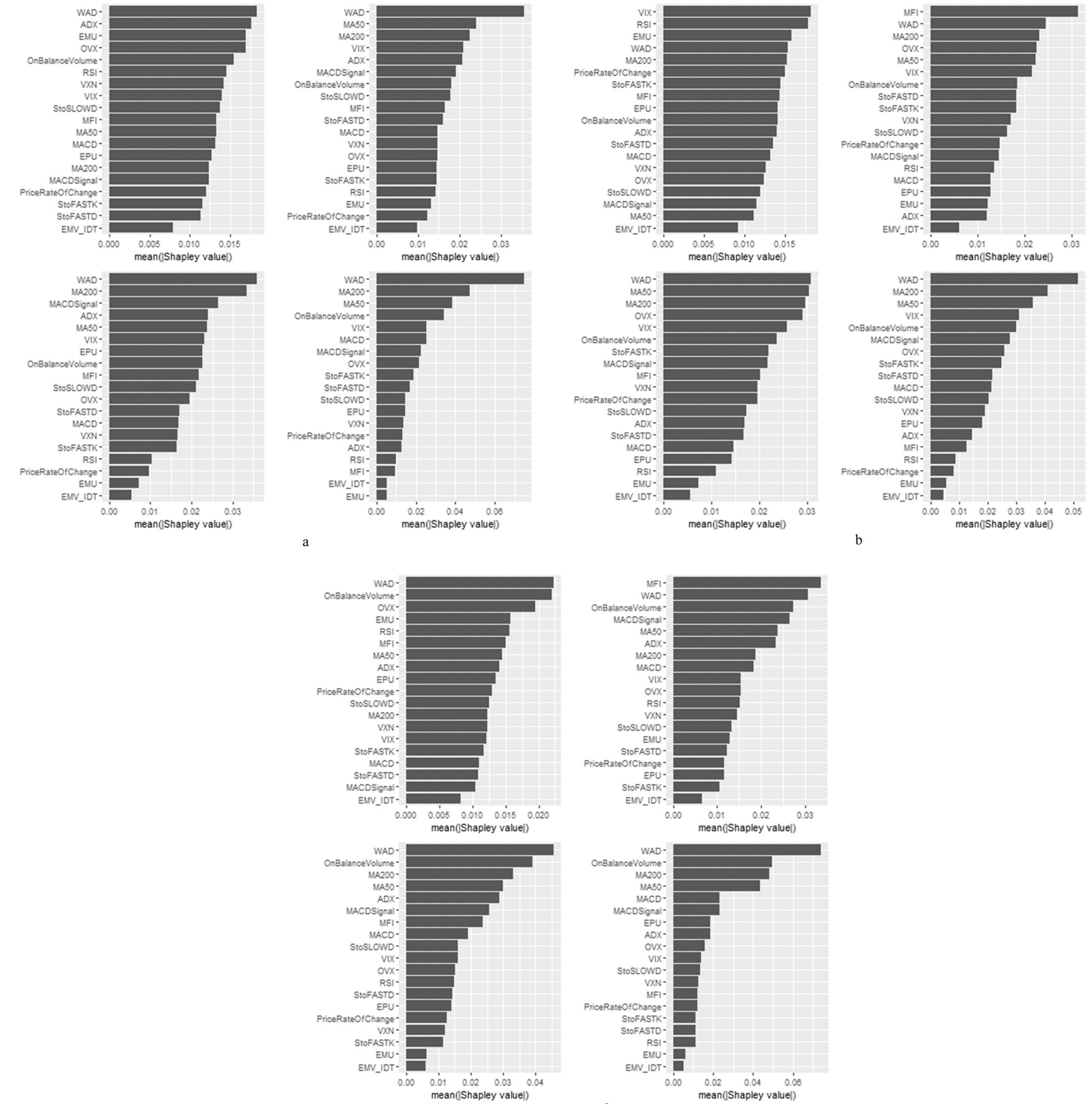

Fig. 6. (a). Shapley values for PBW. Top left (5 day forecast horizon), top right (10 day forecast horizon), bottom left (15 day forecast horizon), and bottom right (20 day forecast horizon). (b). Shapley values for ICLN. Top left (5 day forecast horizon), top right (10 day forecast horizon), bottom left (15 day forecast horizon), and bottom right (20 day forecast horizon). (c). Shapley values for QCLN. Top left (5 day forecast horizon), top right (10 day forecast horizon), bottom left (15 day forecast horizon), and bottom right (20 day forecast horizon).

among the top five most important features while VIX is the most important non-technical feature. Focusing on the 20-day forecasts (because from Fig. 3 forecast accuracy is highest for the 15–20-day forecasts), shows that the top four important features are technical indicators (WAD, MA200, MA50, OnBalanceVolume). The most important non-technical feature is VIX followed by OVX.

For the case of ICLN, WAD, MA200, and MA50 are among the top five features for forecasting 10, 15, and 20 days (Fig. 6b). For 20-day forecasts, the top five features in terms of importance are the same as the those for PBW (WAD, MA200, MA50, VIX, OnBalanceVolume).

For QCLN, the most important features for predicting 15 and 20 days are WAD, OnBalanceVolume, MA200, and MA50 (Fig. 6c). For

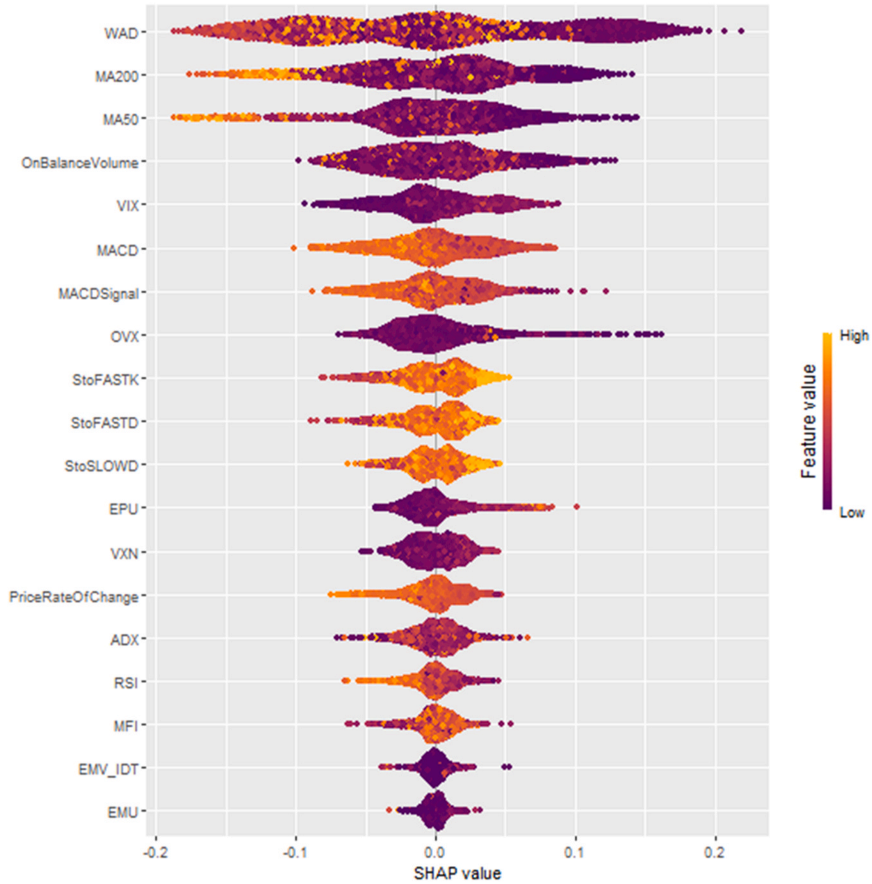

Fig. 7. PBW SHAP variable importance (random forest 20-day forecast).

20-day forecasts, the most important non-technical feature is EPU which is different from PBW and ICLN where EPU ranks below VIX and OVX in terms of importance. The concentration of top ten companies and composition of QCLN is slightly different (US small cap, mid cap, and large cap companies) from that of either PBW or ICLN and this may be the reason for the higher importance of EPU for forecasting QCLN.

The Figs. 6a, 6b, and 6c show variable importance calculated using the mean of the absolute Shapley value but these figures do not show the relationship between the sign of the feature values and importance. A bee swarm plot for PBW 20-day forecast SHAP (Shapley additive explanations) variable importance is also revealing (Fig. 7). The plot is organized so that the most important features are listed in order from the top to the bottom. The ranking of each feature in terms of importance in Fig. 7 is the same as that in Fig. 6a (bottom right) but Fig. 7 includes additional information on the relationship between the importance and feature value. Fig. 7 is color coded so that high feature values have warmer colors and low feature values have cooler colors. The most important feature is WAD and lower values of WAD have a greater importance. It is also the case that lower values for MA200, MA50 and OnBalanceVolume have higher importance. The StoFASTK, StoFASTD, and StoSLOWD are examples of features where higher feature values are associated with greater importance.

In the case of ICLN and QCLN, lower values of WAD, MA200, and MA50 have greater importance (Figs. 8 and 9). Notice that OnBalanceVolume is more difficult to interpret. For PBW and QCLN, lower values are more important while for ICLN higher value are more important. For each of PBW, ICLN, and QCLN, higher values of VIX tend to have more importance than lower values (Figs. 7 – 9). This is also the case for OVX but there is not much variation in the OVX values. These results on VIX are supportive of the findings by Chakrabarti and Sen (2021) who find that green stocks offer protection against downside risks.

The analysis reported so far has relied on cross-validation (CV) where the data is randomly split into a training data set (70% of the data) and a test data set (30% of the data). This approach may be problematic for time series data unless the time series properties of the data are preserved. This is not likely a problem for the applicating in this paper because the models used in this paper use a classification dependent variable with technical indicators as features which contain important time series information. An alternative approach is to use time-series cross-validation (tsCV). With tsCV, the training set occurs immediately prior to the test set. A fixed rolling window approach was used where the length of the rolling window was fixed to use 70% of the data. The random forests model is selected because it has high accuracy over all forecast horizons and is easy to use to calculate variable importance.

In comparing the prediction accuracy between the two methods notice how similar they are (Table 2). For PBW, CV accuracy is 0.8992 while for tsCV the accuracy is 0.9020. For ICLN, CV accuracy is 0.9012

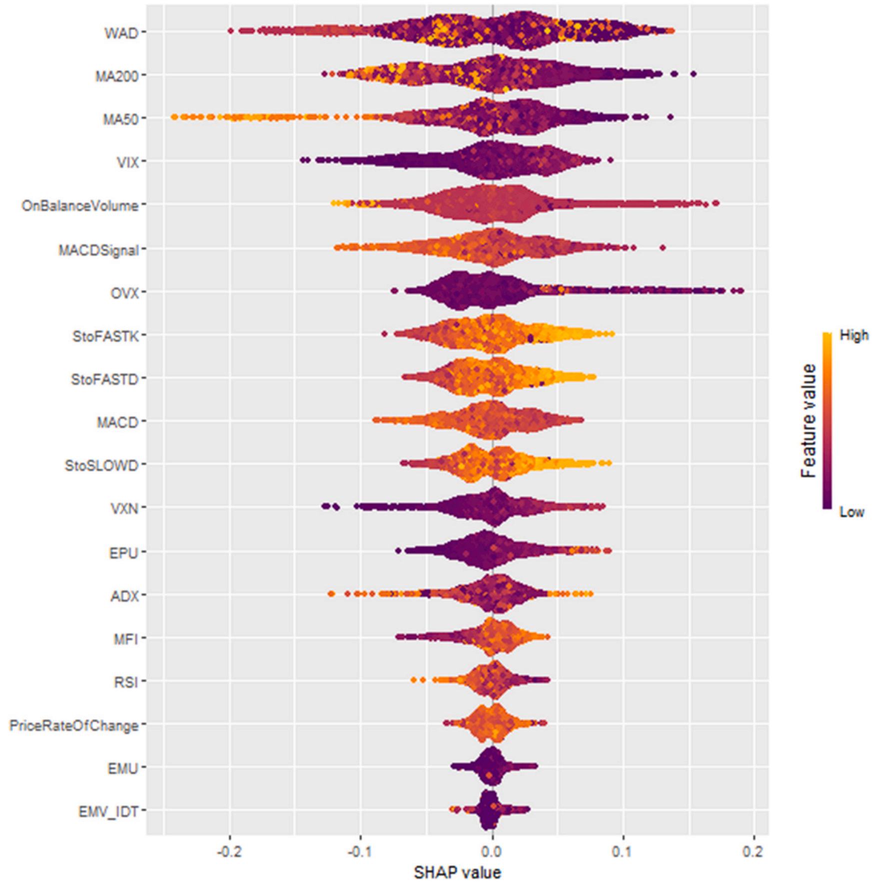

Fig. 8. ICLN SHAP variable importance (random forest 20 -day forecast).

while for tsCV it is 0.9040. For QCLN, CV accuracy is 0.8901 while for tsCV it is 0.8737. For each of PBW, ICLN, and QCLN, the CV and tsCV approaches yield similar values for predictive accuracy, kappa, and FI. Although not reported, the CV and tsCV approaches yield similar values for predictive accuracy for the Extra Trees. The results reported in Table 2 are important in demonstrating that the prediction accuracy measures are robust to the choice of data splitting (CV or tsCV).

In summarizing the results, it can be concluded that for each clean energy ETF, Shapley feature importance indicates that technical indicators like WAD, MA50, and MA200 are consistently ranked among the most important features for the 15-day and 20-day forecast horizons. These results on the importance of technical indicators for predicting clean energy stock prices are consistent with some recent research that uses machine learning methods to predict clean energy stock prices. For example, Sadorsky (2021a) finds that technical indicators are important features for predicting clean energy ETF price direction but he did not include non-technical features. Sadorsky (2022) finds that technical indicators are the most important features for predicting solar stock price direction.

Among the non-technical indicators, VIX and OVX are important features for the 15 and 20 day forecast horizons. The VIX measures the volatility of the S&P 500 and it is not surprising that VIX is an important feature for predicting clean energy ETFs since market volatility can have a pronounced impact on asset prices. The importance of OVX for predicting clean energy stock prices is consistent with recent research demonstrating that oil price volatility has been identified as an important leading economic indicator (Chatziantoniou et al., 2021). The EPU is the most important non-technical feature for predicting 20-day QCLN. This may be due to QCLN having a high concentration (56% of the total assets are accounted for by the top 10 stocks) and its focus on US companies that makes it more susceptive to changes in economic policy uncertainty. Perhaps somewhat surprising is that technology stock market volatility (VXN) is not a very important feature for predicting clean energy stock price direction. Historically, one of the most cited criticisms of the adoption of clean energy is that it is too expensive and very much dependent upon technological innovation. But things have changed. The levelized cost of clean energy like solar and on shore wind has fallen dramatically since 2010 and these clean energy sources are now less expensive than that of coal (The Economist, 2020). Clean energy may have experienced the technological revolution needed to hasten widespread adoption.

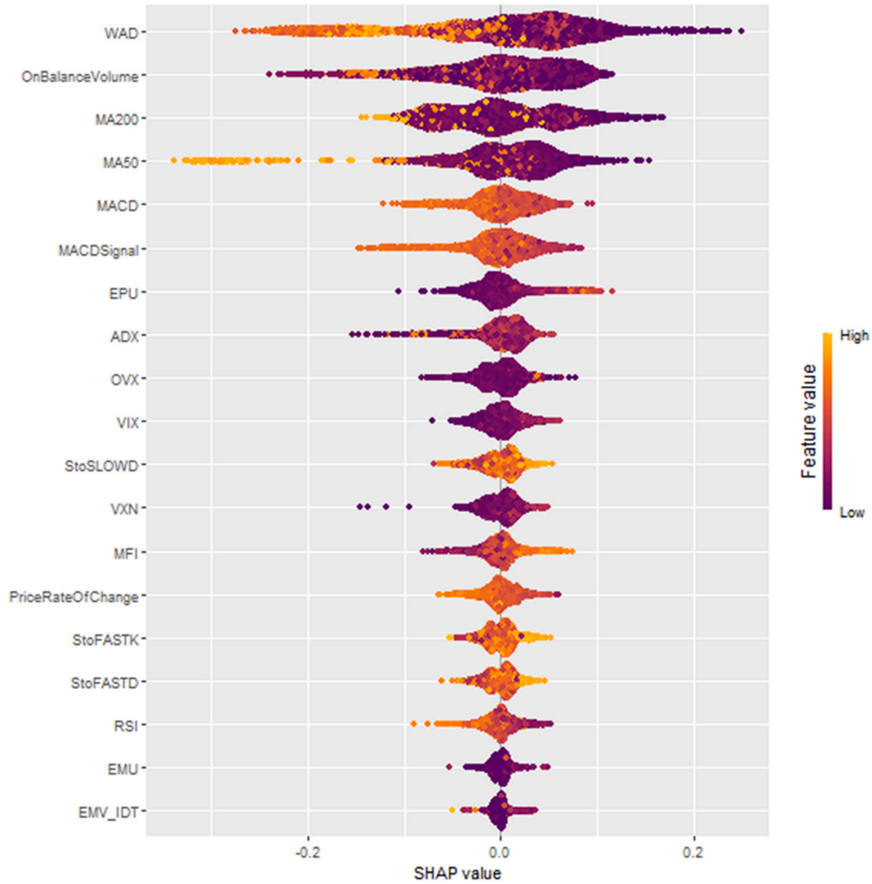

Fig. 9. QCLN SHAP variable importance (random forest 20-day forecast).

Table 2
Prediction accuracy for random forests calculated from CV and tsCV.

	20-day CV	20-day tsCV
PBW		
Accuracy	0.8992	0.9020
Kappa	0.7972	0.7960
F1	0.9064	0.8777
ICLN		
Accuracy	0.9012	0.9040
Kappa	0.8009	0.7949
F1	0.9093	0.8714
QCLN		
Accuracy	0.8901	0.8737
Kappa	0.7714	0.7279
F1	0.9083	0.8271

6. Conclusions and practical implications

Climate change is one of the biggest risks facing society. Any serious efforts to addressing climate change requires fuel switching from fossil fuels to renewables. Thus, clean energy has a pivotal role to play in the transition to a more sustainable economy. Encouragingly, clean energy is, currently, the fastest growing energy source. This growth in clean energy is expected to continue as countries transition to lower carbon energy sources. Clean energy equities are growing in popularity and investors need accurate clean energy stock price forecasts in order to make sound investment decisions. Surprisingly, little work has been done on the topic of forecasting clean energy stock prices. The aim of this study is to identify how important volatility and uncertainty are for forecasting clean energy stock price direction. Clean energy stock prices are measured using three widely traded ETFs (PBW, ICLN, QCLN).

The introduction to this paper posed three important questions regarding forecasting clean energy equity prices. In terms of the first question, “which forecasting method to use?”, this research finds random forests, stochastic gradient boosting, and extremely randomized trees produce higher accuracy forecasts than those from Lasso or Naïve Bayes models. Random forests, GBM, and extremely randomized trees have prediction accuracy greater than 80% after eight days. After fifteen days, the forecast accuracy from these models is greater than 85%. In some cases, prediction accuracy is an impressive 90% or even slightly higher. SVMs have high prediction accuracy but not as high as the predictions achieved using random forests, GBM, and extremely randomized trees. Extra Trees have the highest prediction accuracy for most forecast horizons. By comparison, prediction accuracy from Lasso is never higher than 65%. The prediction accuracy of Naïve Bayes is mostly less than that of Lasso. The results from this research add support to the literature demonstrating the accuracy of predicting the direction of stock prices with machine learning methods.

In answering the question of how important market volatility and economic policy uncertainty are for predicting clean energy stock prices, the analysis reveals that the most important features for predicting the direction of clean energy stock prices are the technical indicators WAD, MA50, and MA200. These results are consistent with the literature that establishes technical indicators as important features for predicting stock prices. In most cases, the most important non-technical features for the 15–20 day forecast period are VIX and OVX. Unlike in the case of PBW or ICLN, economic policy uncertainty is important for predicting 20-day QCLN. In other cases, economic policy uncertainty is not that important. Text based equity market uncertainty is an important predictor for making five day forecasts but otherwise one of the least important features. Somewhat surprising is that technology stock market volatility is not a very important predictor of clean energy stock prices. This may be due to the dramatic drop in costs of supplying clean energy which lessens the reliance clean energy development on technological innovation.

With regards to the question as to whether variable importance is consistent across different forecast horizons, the answer is mixed. Variable importance tends to vary little between the 15 day and 20 day forecast horizons. This is relevant since the highest prediction accuracy is observed between this range of days. Variable importance does change between five days and ten days and is most noticeable when comparing the 5 day forecasts with the 20 day forecasts but forecast accuracy is lowest for the one to five day forecast horizon.

This research offers several practical implications. First, random forests, stochastic gradient boosting, and extremely randomized trees should be used to forecast clean energy stock price direction because the predictions from these methods are more accurate than those computed from Lasso and Naïve Bayes models. Random forests and extremely randomized trees are especially worth using because they are easy to estimate and take relatively little computation time. One recommendation is to use extremely randomized trees to forecast clean energy stock price directions because this method is computationally fast and produces high accuracy. If one is interested in high forecast accuracy and a closer examination of variable importance than random forests are a good choice.

Second, the features used for clean energy stock price prediction should include technical indicators (like WAD and moving averages), stock market volatility (VIX) and oil price volatility (OVX) because these features rank high in variable importance across a range of forecast periods. The literature on clean energy stock prices has established that oil is an important variable impacting clean energy stock prices and the results from this paper demonstrate the importance of the volatility of oil prices as an important predictor of clean energy stock price direction.

There are several avenues for future research. One avenue for future research would be to expand the predictor space to include interest rate and exchange rate variables. Including these variables may offer a more complete picture of how business cycle conditions affect clean energy stock price predictability. Another possible avenue for future research would be to expand the number of methods used. The forecasting accuracy of some of the models used in this paper are impressively high in the fifteen day to twenty day forecast range but it may be interesting to see if additional methods like deep learning or ensemble stacking can improve forecast accuracy over shorter forecast periods.

Acknowledgements

I would like to thank participants at the Ethical Finance and Sustainable Conference: Energy Transition & Sustainability (EDC Paris Business School, Paris, May 12-13, 2022) for helpful comments and in particular Zied Ftiti for inviting me. I thank an anonymous reviewer for helpful comments. I have benefited from conversations with Irene Henriques, Duc Khuong Nguyen, Ugur Soytas, Sami Ben Jabeur, and Muhammad Ali Nasir.

References

Ahmad, W., Sadorsky, P., & Sharma, A. (2018). Optimal hedge ratios for clean energy equities. *Economic Modelling, 72*, 278–295. https://doi.org/10.1016/j.econmod.2018.02.008

Ampomah, E. K., Qin, Z., & Nyame, G. (2020). Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. *Information, 11*, 332. https://doi.org/10.3390/info11060332

Arouri, M., Estay, C., Rault, C., & Roubaud, D. (2016). Economic policy uncertainty and stock markets: Long-run evidence from the US. *Finance Research Letters, 18*, 136–141. https://doi.org/10.1016/j.frl.2016.04.011

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics, 131*, 1593–1636. https://doi.org/10.1093/qje/qjw024

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications, 42*, 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance, 47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013

BloombergNEF, 2022. Energy Transition Investment Trends 2022.

Bondia, R., Ghosh, S., & Kanjilal, K. (2016). International crude oil prices and the stock prices of clean energy and technology companies: Evidence from non-linear cointegration tests with unknown structural breaks. *Energy, 101*, 558–565. https://doi.org/10.1016/j.energy.2016.02.031

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications, 156*, Article 113464. https://doi.org/10.1016/j.eswa.2020.113464

Chakrabarti, G., & Sen, C. (2021). Dynamic market risk of green stocks across regions: Where does the devil lie. *Journal of Cleaner Production, 303*, Article 127028. https://doi.org/10.1016/j.jclepro.2021.127028

Chatziantoniou, I., Filippidis, M., Filis, G., & Gabauer, D. (2021). A closer look into the global determinants of oil price volatility. *Energy Economics, 95*, Article 105092. https://doi.org/10.1016/j.eneco.2020.105092

Dutta, A. (2019). Impact of silver price uncertainty on solar energy firms. *Journal of Cleaner Production, 225*, 1044–1051. https://doi.org/10.1016/j.jclepro.2019.04.040

Dutta, A. (2018). Oil and energy sector stock markets: An analysis of implied volatility indexes. *Journal of Multinational Financial Management, 44*, 61–68. https://doi.org/10.1016/j.mulfin.2017.12.002

Dutta, A. (2017). Oil price uncertainty and clean energy stock returns: New evidence from crude oil volatility index. *Journal of Cleaner Production, 164*, 1157–1166. https://doi.org/10.1016/j.jclepro.2017.07.050

Dutta, A., Bouri, E., Das, D., & Roubaud, D. (2020). Assessment and optimization of clean energy equity risks and commodity price volatility indexes: Implications for sustainability. *Journal of Cleaner Production, 243*, Article 118669. https://doi.org/10.1016/j.jclepro.2019.118669

Dutta, A., Bouri, E., & Noor, M. H. (2018). Return and volatility linkages between CO_2 emission and clean energy stock prices. *Energy, 164*, 803–810.

Elie, B., Naji, J., Dutta, A., & Uddin, G. S. (2019a). Gold and crude oil as safe-haven assets for clean energy stock indices: Blended copulas approach. *Energy, 178*, 544–553.

Elie, B., Naji, J., Dutta, A., & Uddin, G. S. (2019b). Gold and crude oil as safe-haven assets for clean energy stock indices: Blended copulas approach. *Energy, 178*, 544–553. https://doi.org/10.1016/j.energy.2019.04.155

Ferrer, R., Shahzad, S. J. H., López, R., & Jareño, F. (2018). Time and frequency dynamics of connectedness between renewable energy stocks and crude oil prices. *Energy Economics, 76*, 1–20. https://doi.org/10.1016/j.eneco.2018.09.022

Geng, J.-B., Liu, C., Ji, Q., & Zhang, D. (2021). Do oil price changes really matter for clean energy returns. *Renewable and Sustainable Energy Reviews, 150*, Article 111429. https://doi.org/10.1016/j.rser.2021.111429

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Ghosh, P., Neufeld, A., & Sahoo, J. K. (2021). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*Article 102280. https://doi.org/10.1016/j.frl.2021.102280

Greenwell, B. , (2021). fastshap: Fast Approximate Shapley Values.

Greenwell, B. , Boehmke, B. , Cunningham, J. , Developers (〈https://github.com/gbm-developers〉), G.B.M., (2020). gbm: Generalized Boosted Regression Models.

Gupta, K. (2017). Do economic and societal factors influence the financial performance of alternative energy firms. *Energy Economics, 65*, 172–182. https://doi.org/10.1016/j.eneco.2017.05.004

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer Series in Statistics. Springer-Verlag,. 〈https://doi.org/10.1007/978-0-387-84858-7〉.

Henriques, I., & Sadorsky, P. (2008). Oil prices and the stock prices of alternative energy companies. *Energy Economics, 30*, 998–1010. https://doi.org/10.1016/j.eneco.2007.11.001

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R, Springer Texts in Statistics*. New York: Springer-Verlag,https://doi.org/10.1007/978-1-4614-7138-7

Ji, Q., Liu, B.-Y., Nehler, H., & Uddin, G. S. (2018). Uncertainties and extreme risk spillover in the energy markets: A time-varying copula-based CoVaR approach. *Energy Economics, 76*, 115–126. https://doi.org/10.1016/j.eneco.2018.10.010

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-01839-w

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2020. caret: Classification and Regression Training.

Kumar, S., Managi, S., & Matsuda, A. (2012). Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics, 34*, 215–226. https://doi.org/10.1016/j.eneco.2011.03.002

Le, T.-L., Abakah, E. J. A., & Tiwari, A. K. (2021). Time and frequency domain connectedness and spill-over among fintech, green bonds and cryptocurrencies in the age of the fourth industrial revolution. *Technological Forecasting and Social Change, 162*, Article 120382. https://doi.org/10.1016/j.techfore.2020.120382

Leung, M. T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting, 16*, 173–190. https://doi.org/10.1016/S0169-2070(99)00048-5

Liu, T., & Hamori, S. (2020). Spillovers to renewable energy stocks in the US and Europe: Are they different. *Energies, 13*, 3162. https://doi.org/10.3390/en13123162

Liu, T., Nakajima, T., & Hamori, S. (2022). The impact of economic uncertainty caused by COVID-19 on renewable energy stocks. *Empirical Economics, 62*, 1495–1515. https://doi.org/10.1007/s00181-021-02087-3

Lohrmann, C., & Luukka, P. (2019). Classification of intraday S&P500 returns with a random forest. *International Journal of Forecasting, Special Section: Supply Chain Forecasting, 35*, 390–407. https://doi.org/10.1016/j.ijforecast.2018.08.004

Lundberg, S., Lee, S.-I., 2016. An unexpected unity among methods for interpreting model predictions. Presented at the Advances in Neural Information Processing Systems, arXiv. https://doi.org/10.48550/arXiv.1611.07478.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*. Curran Associates, Inc..

Lundgren, A. I., Milicevic, A., Uddin, G. S., & Kang, S. H. (2018). Connectedness network and dependence structure mechanism in green investments. *Energy Economics, 72*, 145–153. https://doi.org/10.1016/j.eneco.2018.04.015

Maghyereh, A. I., Awartani, B., & Abdoh, H. (2019). The co-movement between oil and clean energy stocks: A wavelet-based analysis of horizon associations. *Energy, 169*, 895–913.

Managi, S., & Okimoto, T. (2013). Does the price of oil interact with clean energy prices in the stock market. *Japan and the World Economy, 27*, 1–9. https://doi.org/10.1016/j.japwor.2013.03.003

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., C++-code), C.-C.C. (libsvm, C++-code), C.-C.L. (libsvm, 2021. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*, 87–106. https://doi.org/10.1257/jep.31.2.87

Naeem, M. A., Peng, Z., Suleman, M. T., Nepal, R., & Shahzad, S. J. H. (2020). Time and frequency connectedness among oil shocks, electricity and clean energy markets. *Energy Economics, 91*, Article 104914. https://doi.org/10.1016/j.eneco.2020.104914

Nasreen, S., Tiwari, A. K., Eizaguirre, J. C., & Wohar, M. E. (2020). Dynamic connectedness between oil prices and stock returns of clean energy and technology companies. *Journal of Cleaner Production, 260*, Article 121015. https://doi.org/10.1016/j.jclepro.2020.121015

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science, 60*, 1772–1791. https://doi.org/10.1287/mnsc.2013.1838

Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting, 27*, 561–578. https://doi.org/10.1016/j.ijforecast.2010.02.008

Nyberg, H., & Pönkä, H. (2016). International sign predictability of stock returns: The role of the United States. *Economic Modelling, 58*, 323–338. https://doi.org/10.1016/j.econmod.2016.06.013

Pástor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance, 67*, 1219–1264. https://doi.org/10.1111/j.1540-6261.2012.01746.x

Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance, 9*, 495–510. https://doi.org/10.1016/S0927-5398(02)00007-5

Pham, L. (2021). Frequency connectedness and cross-quantile dependence between green bond and green equity markets. *Energy Economics, 98*, Article 105257. https://doi.org/10.1016/j.eneco.2021.105257

Pham, L. (2019). Do all clean energy stocks respond homogeneously to oil price. *Energy Economics, 81*, 355–379. https://doi.org/10.1016/j.eneco.2019.04.010

Pönkä, H. (2016). Real oil prices and the international sign predictability of stock returns. *Finance Research Letters, 17*, 79–87. https://doi.org/10.1016/j.frl.2016.01.011

R Core Team (2022). *R: A language and environment for statistical computing.* Vienna, Austria: The R Project for Statistical Computing,.

Reboredo, J. C. (2015). Is there dependence and systemic risk between oil and renewable energy stock prices. *Energy Economics, 48*, 32–45. https://doi.org/10.1016/j.eneco.2014.12.009

Reboredo, J. C., Quintela, M., & Otero, L. A. (2017a). Do investors pay a premium for going green? Evidence from alternative energy mutual funds. *Renewable and Sustainable Energy Reviews, 73*, 512–520. https://doi.org/10.1016/j.rser.2017.01.158

Reboredo, J. C., Rivera-Castro, M. A., & Ugolini, A. (2017b). Wavelet-based test of co-movement and causality between oil and renewable energy stock prices. *Energy Economics, 61*, 241–252. https://doi.org/10.1016/j.eneco.2016.10.015

Reboredo, J. C., & Ugolini, A. (2018). The impact of energy prices on clean energy stock prices. A multivariate quantile dependence approach. *Energy Economics, 76*, 136–152. https://doi.org/10.1016/j.eneco.2018.10.012

Sadorsky, P. (2022). Forecasting solar stock prices using tree-based machine learning classification: How important are silver prices. *The North American Journal of Economics and Finance, 61*, Article 101705. https://doi.org/10.1016/j.najef.2022.101705

Sadorsky, P. (2021a). A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management, 14*, 48. https://doi.org/10.3390/jrfm14020048

Sadorsky, P. (2021b). Predicting gold and silver price direction using tree-based classifiers. *Journal of Risk and Financial Management, 14*, 198. https://doi.org/10.3390/jrfm14050198

Saeed, T., Bouri, E., & Alsulami, H. (2021). Extreme return connectedness and its determinants between clean/green and dirty energy investments. *Energy Economics, 96*, Article 105017. https://doi.org/10.1016/j.eneco.2020.105017

Schwert, G. W. (1990). Stock market volatility. *Financial Analysts Journal, 46*, 23–34. https://doi.org/10.2469/faj.v46.n3.23

Simm, J., Abril, I.M. de, 2014. extraTrees: Extremely Randomized Trees (ExtraTrees) Method for Classification and Regression.

The Economist, 2020. The world's energy system must be transformed completely. The Economist May 23 edition.

Uddin, G. S., Rahman, M. L., Hedström, A., & Ahmed, A. (2019). Cross-quantilogram-based correlation and dependence between renewable energy stock and other asset classes. *Energy Economics, 80*, 743–759. https://doi.org/10.1016/j.eneco.2019.02.014

Ulrich, J., 2020. TTR: Technical Trading Rules.

Wang, Y., Liu, L., & Wu, C. (2020). Forecasting commodity prices out-of-sample: Can technical indicators help. *International Journal of Forecasting, 36*, 666–683. https://doi.org/10.1016/j.ijforecast.2019.08.004

Wen, X., Guo, Y., Wei, Y., & Huang, D. (2014). How do the stock prices of new energy and fossil fuel companies correlate? Evidence from China. *Energy Economics, 41*, 63–75. https://doi.org/10.1016/j.eneco.2013.10.018

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications, 112*, 258–273. https://doi.org/10.1016/j.eswa.2018.06.016

Wright, M.N., Wager, S., Probst, P., 2022. ranger: A Fast Implementation of Random Forests.

Yin, L., & Yang, Q. (2016). Predicting the oil prices: Do technical indicators help. *Energy Economics, 56*, 338–350. https://doi.org/10.1016/j.eneco.2016.03.017

Yin, L., Yang, Q., & Su, Z. (2017). Predictability of structural co-movement in commodity prices: the role of technical indicators. *Quantitative Finance, 17*, 795–812. https://doi.org/10.1080/14697688.2016.1225977

Zhao, X. (2020). Do the stock returns of clean energy corporations respond to oil price shocks and policy uncertainty. *Economic Structures, 9*, 53. https://doi.org/10.1186/s40008-020-00229-x