Energy Economics 101 (2021) 105393

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

Influences of sentiment from news articles on EU carbon prices

Jing Ye, Minggao Xue *

School of Management, Huazhong University of Science and Technology, Wuhan 430070, China

A R T I C L E I N F O

Keywords:
Textual analysis
Carbon pricing
Carbon tone index
LDA
Carbon price prediction

A B S T R A C T

In this paper, we calculate the carbon tone index that reflects sentiment in news articles through a self-built dictionary. We study the effect of carbon tone index on carbon price return in the period from September 19, 2017 to October 9, 2020. In addition, we employ the Latent Dirichlet Allocation (LDA) method to explore the differential influences of different topic carbon tone indexes on carbon prices. The market confidence boosted by the approved Market Stability Reserve (MSR) policy led to a continuous increase in volume in 2018. Using two subsample periods divided by the implementation of MSR, we explore the problem whether the increased high volume changed the speed of information absorption in carbon market. Quantile regression with control variables (coal, oil, natural gas, electricity and stock prices) is used to test the robustness of the estimated results. The empirical results show that carbon tone index is closely associated with changes in carbon prices and the efficiency of carbon market is improved after MSR. Finally, we use all carbon tone indexes at the 10% significance level in eight predictive models and show the economic value of the optimal predictive model. In summary, carbon tone index has a strong predictive power for carbon prices.

1. Introduction

Carbon pricing is a market-based strategy for mitigating global warming. The European Union Emission Trading System (EU ETS) is currently the world's largest carbon emission trading market with more than three-quarters of the total international carbon trading volume, and has a demonstrative effect on other carbon markets. EU ETS had maintained a low price in the early stage of the phase III and did not start a round of steep rising until 2017. Unexpected, the continuous and large changes in carbon prices couldn't be entirely attributed to fundamentals, because influencing factors did not undergo the corresponding changes during this period. The sentiment expressed in the news could convey information and opinions about many aspects of the market to the market participants (Kearney and Liu, 2014). Tetlock (2007) confirmed that the news tone affected subsequent changes in stock prices. We attempt to use carbon tone index extracted from news articles for carbon pricing in this paper. In addition, some scholars believe that the speed at which market participants incorporate information into asset prices is associated with market efficiency. The slow dissemination of public information in asset prices means low efficiency in the market (Duffie, 2010b; Li, 2006; Loughran et al., 2019). In this paper, according to the speed of the carbon market's response to news information, the efficiency of carbon market will be indirectly explored.

We use a self-built word list to measure carbon tone index in news articles. We abandon popular machine learning methods due to the difficulty to achieve reproducibility (Loughran and Mcdonald, 2016) and the time-consuming manual tagging (Kearney and Liu, 2014). The other point is that, as everyone knows, big data is the cornerstone of machine learning, but carbon market as an emerging market has insufficient news articles in the early stage of market establishment. Following the method proposed by Loughran et al. (2019), we randomly select a large number of samples from news articles, and manually extract keywords that convey information signals. Compared with classic General Inquirer (GI) built-in dictionary (Stone et al., 1966), Harvard IV-4 dictionary[1] and L&M finance dictionary (Loughran and Mcdonald, 2011), our word list is small with a total of 198 keywords. This article, as a pioneering study in carbon finance, extracts sentiment index from news articles about EU ETS. We carefully select keywords by emphasizing the signal-to-noise ratio (Loughran and Mcdonald, 2016). Shapiro et al. (2020) also indicated that the domain-specific word list was conducive to extracting more accurate information. For example, the EU ETS is a cap-and-trade system dominated by the government. The government decisions involving the supply and demand of allowances influence carbon prices, such as supporting stricter emission reduction

* Corresponding author.
E-mail addresses: yejing24@hust.edu.cn (J. Ye), xmgwt@mail.hust.edu.cn (M. Xue).

[1] The dictionary is offered by the website http://www.wjh.harvard.edu/~inquirer/3JMoreInfo.html.

https://doi.org/10.1016/j.eneco.2021.105393
Received 22 March 2021; Received in revised form 31 May 2021; Accepted 10 June 2021
Available online 15 June 2021
0140-9883/

long-term targets and higher linear reduction factors. The frequent and professional keywords like approval, support, exit and Brexit may capture the impact of government decisions on carbon prices. We first use ordinary least squares (OLS) regression to test the effect of carbon tone index on carbon prices without any control variables as an initial step.

In this paper, the Latent Dirichlet Allocation (LDA) method is used to answer the question of how different news topic content affects carbon prices. Brandt and Gao (2019) indicated that the text analysis technologies provided a quantitative method for the effects that different events caused. We match the classification labels with the realistic topics (or event types) based on these topic words obtained by LDA method. The news topic information could increase the interpretation of asset price changes (Brandt and Gao, 2019; Caporin and Poli, 2017; Deveikyte et al., 2020; Nguyen et al., 2015). Specifically, Brandt and Gao (2019) found that macroeconomic news and geopolitical events had different impact patterns and predictive abilities for crude oil prices. For EU ETS, we observe various news topics such as some regular allowances auction, irregular discussions and voting from European Parliament. Tetlock et al., 2008 hold that analyzing the complete set of events that affect asset price changes could help governments and investors understand the reaction patterns of events for policy planning and investment decisions. Hence, we analyze the impact of topic carbon tone index on the formation of carbon prices in OLS model.

Some scholars found that sentiment exerted different effects on asset prices in different market environments (Zou and Sun, 2012). Using quantile regression model with control variables, we robustly test the response of carbon prices to carbon tone index at different quantiles. The quantile regression obtains the estimated coefficients by minimizing the sum of absolute weighted errors, and the estimator is robust and not easily affected by outliers. The quantile regression estimator may be more effective than the least square estimator in the case of non-normal distribution of disturbance term. The set of control variables includes coal, oil, natural gas, electricity and stock prices (Aatola et al., 2013; Alberola et al., 2008; Chevallier, 2011; Dutta et al., 2018; Koch et al., 2014). In addition, we also include the lagged carbon price returns determined by partial autocorrelation function. Limited by data availability, our sample is collected from September 19, 2017 to October 9, 2020. According to a simple statistical analysis, we find a continuously high volume in EU ETS before the implementation of Market Stability Reserve (MSR). Consistently, the Refinitiv report in 2018 roughly estimated that the total volume of allowances declined about 4% in 2019 after jumping 45% in 2018.[2] The key performance indicators for evaluating carbon market functions show increased liquidity, volume, auction coverage and open positions in EU ETS in 2018 (Marcu et al., 2019). Marcu et al. (2019) also indicated that the increased financial participants brought a large number of transactions. In view of the more stringent climate policy approved by the carbon trading system reform in Phase IV, financial investment institutions participated in carbon market to obtain benefits from high carbon prices in the future. We expect that the high volume that the diversified market participants and higher open positions caused may change market attention and hence influence the market's response structure to information. For example, the high open positions mean high risks, which may increase the market attention of investment institutions. Therefore, investors will focus on the market information in real time and adjust their positions according to the specific strategies. Investor attention also will increase to make market more efficient. Empirical researches indicated that investor attention could improve the ability of price discovery and reduce price distortions caused by information asymmetry (Andrei and Hasler, 2014). We divide the full sample into two subsamples, and then verify whether the speed of information assimilation has changed after MSR.

In the empirical results, we find that carbon tone index has a significant impact on carbon price. Importantly, carbon tone index slowly acted on carbon market in the subsample period from September 19, 2017 to January 1, 2019. After a continuous increase in trading volume, carbon market quickly showed a negative overreaction to carbon tone index in the subsample period from January 1, 2019 to October 9, 2020. We obtain the same findings robustly in quantile regression. In addition, carbon tone index has a higher coefficient that acts on carbon price return at the two-sides tail quantiles. With the contribution of LDA model, we find that carbon tone index based on news about the supply and demand of allowances has the most robust impact on changes in carbon price. Topic carbon tone index based on policy-making in EU ETS produces a positive 0.2807 effect at the 10% significance level. Differently, carbon tone index involving allowances auctions slowly shows a negative impact on carbon price return. Finally, we also implement an application that using these carbon tone indexes for carbon price prediction. We include all carbon tone indexes at the 10% significant level in six machine learning models. We set up two experimental groups: including only fundamental factors and including additionally carbon tone indexes. The results show that the carbon tone indexes have a strong prediction ability in any models. We also implement the market timing strategy to obtain an average annualized return of 47.69% by the optimal predictive model (Principal Components Regression (PCR)).

2. Relevant theories and literature

To date, there has been many empirical literature about carbon pricing mechanism of EU ETS based on the marginal abatement theory. An effective carbon market should play a role in discovering the marginal abatement costs of the entire society. Economic activities (Chevallier, 2009; Gronwald et al., 2011), energy prices (including coal, oil and gas) (Hammoudeh et al., 2014; Sousa et al., 2014; Tan and Wang, 2017) and electricity prices (Zhu et al., 2019) as fundamentals were empirically found that had statistically significant effects. The professional EU ETS report believed that the market sentiment was more important than fundamentals to a certain extent (Marcu et al., 2019). The state of the EU ETS report (Marcu et al., 2018; Marcu et al., 2019) and the multinational organizations such as PricewaterhouseCoopers (PwC) (Association, 2012) both designed questionnaires to obtain the carbon market sentiment. Keynes (2018) first proposed the concept of "animal spirit" and hold that investors driven by sentiments would lead to changes in asset prices. However, few researches discussed the effect of sentiment on carbon prices.

Constructing quantitative sentiment indicators may produce barriers to in-depth research. Almost all existing studies used some market activity indicators to represent the carbon market sentiment. Koch et al. (2014) and Jiao et al. (2018) found that the monthly economic sentiment indicators of EU could explain changes in carbon prices. More directly, Deeney et al. (2016) used the principal component method to combine indicators such as volume and open interest as the sentiment level in carbon market, which affects the effect of government announcements on carbon prices. In addition to EU ETS, Fan and Todorova (2017) regarded China's VIX and OVX as sentiment indicators in China's pilot carbon markets. The above indicators are limited by the frequency of collected data and usually cannot directly reflect investor beliefs about carbon market. With the development of information technology, text mining technology is popular in the fields of accounting finance and asset pricing. Some scholars attempted to extract sentiment index from media news articles obtained by the crawler technology. Investors always interpret information unbiasedly in classical theory. However, the behavioral finance believes that investors are not rational but have behavioral deviations. Investors in financial market may be affected by noise (Black, 1986; Long et al., 1990; Mendel and Shleifer, 2012) and cognition bias (Barberis et al., 1998; Daniel et al., 1998). The media sentiment conveyed by news articles resonates with investors, and the resulting opinion climate will trigger investors' behaviors and ultimately affect asset prices (You and Wu, 2012). Shiller (2000) indicated

[2] The report is collected from https://www.refinitiv.com/en/resources/special-report/review-of-carbon-markets-in-2018.

that news media had the ability to influence readers' beliefs by choosing the tone to emphasize specific positive or negative events. Empirically, Tetlock (2007) studied in depth how the sentiment index, extracted from the "Abreast of the Market" column of the Wall Street Journal, affected the subsequent stock returns, and found that news articles could predict changes in activity indicators in stock market. In addition, text analysis techniques are also used to capture the sentiment in earnings announcements (Heston and Sinha, 2016), Merger and Acquisition (Yang et al., 2019), company annual reports (Li, 2006) and initial public offering (IPO) prospectus (Ferris et al., 2012), and the empirical researches associated them with activity indicators such as company earning, stock return and trading volume.

More specifically, the recent literature focused on the differentiated market responses from various topic events. The topic model realizes the classification of media news. We match every category to event topic according to the keywords in topic model. Some studies establish topic sentiment indicators based on different event topics to explore the impact of different events on asset prices. Nguyen et al. (2015) attempted to extract topic sentiment indices and market sentiment index at the same time by the joint sentiment topic model (JST), and found that topic sentiment indices from different event topics had different effects on stock price. Brandt and Gao (2019) believed that news analysis provided a method to quantify macroeconomic events and geopolitical events. It can not only provide real-time news content analysis, but also capture the original information in news and market cognition. Brandt and Gao (2019) used topic sentiment scores that the RavenPack database offered to study the effect of different events on oil prices. Li et al. (2019) used Latent Dirichlet Allocation (LDA) method for distinguishing the influence of various online topic news. The empirical results showed that sentiment indicators effectively helped crude oil pricing.

We review the empirical studies on text sentiment in the financial field. Loughran and Mcdonald (2016) indicated that the sentiment index from text conveyed the incremental information in quantitative financial information and had the ability to predict market trends. More obviously, the sentiment measure based on dictionary includes many words or phrases that indicate the relationship between supply and demand (Loughran et al., 2019), and non-financial performance indicators (Krishnamoorthy, 2018). In this way, we also can explore the market efficiency according to the speed of information absorption. The slow information dissemination in assets prices means low efficiency in market, but immediate dissemination means the well-functioning financial markets (Loughran et al., 2019). Sinha (2016) found that the US stock market had underreaction to the tone of news articles. Loughran et al. (2019) showed that crude oil market overreacted to information that reflected the relationship between supply and demand, and indicated that this result may be caused by the lack of investors' attention (Duffie, 2010a). Increased investor attention can promote information dissemination, that is, the accelerated speed at which information is reflected in stock prices.

In this paper, we use text mining to explore the impact of sentiment from media news articles on changes in carbon price. In addition, we discuss the differential impact of different topic news by LDA model. In particular, it is well known that the structural adjustment measures such as MSR policy restores the confidence of participants in carbon market. EUA price reached 20–25 Euro/ton in 2018, and always remained this carbon price level. The total volume also increased by 42% year-on-year. Much of the volume growth is attributable to the return of financial participants, who predict the impact of expected allowance shortages in carbon market (Marcu et al., 2019). We expect that changes in the speed of information absorption will appear after MSR policy due to the increase in market attention. None of the above issues have been explored so far, so our research is a new beginning.

3. Data and method

3.1. Data source

We crawl daily carbon-related news in EU ETS from https://vertis.com and https://carbon-pulse.com. The former cooperates with more than 800 industrial customers and 300 aircraft operators in the power generation and industrial sectors throughout the EU, while it helps them buy and sell different emission trading units and manages their risks. We select the news columns including technical and fundamental analysis from the vertis website. The latter is established by three ex-Reuters/ Point Carbon journalists with almost 30 years' experience in covering carbon markets and climate policy. Carbon Pulse provides in-depth news and intelligence about global carbon pricing schemes and climate change policies. We crawl only all newsletter in EU ETS section under EMEA column from the Carbon Pulse website due to the paid full text service. The newsletter contains sufficient key information and has less noise information in general. Based on the behavioral habits, investors usually obtain the real-time message in the form of email, which offers daily newsletter. We also check other news resources covering carbon market news but they are classified under the energy column or green finance column and publish only little news that closely related to the carbon market. Accessing directly to the relevant news will easily generate a lot of noise. Therefore, considering relevance and feasibility, we choose only vertis full-text news and carbon-pulse newsletter.

Data is collected from September 19, 2017 to October 9, 2020. The sample period is determined by the data availability of control variables and the start time of research. We divide news at the interval of one day. During the full sample period, we crawl a total of 2040 valid news, with an average of nearly 3 news per trading day. The number of daily news releases in carbon market is comparable to that in crude oil market (Loughran et al., 2019). The number of monthly carbon news articles is showed in Fig. 1. From 2018 to 2020, the total number of monthly news showed an upward trend year by year. We use daily settlement prices of continuous European Union Allowances (EUA) futures from Intercontinental Exchange (ICE). We choose control variables including NBP Natural Gas Future (Gas) , Phelix Baseload Future (Electricity) , Rotterdam Coal Future (Coal) , STOXX Europe 600 Index (Stoxx) , Brent Oil Future (Oil) and European Renewable Energy Total Return Index (ERIX). We exclude ERIX, which has a high correlation with Stoxx. All daily variables except carbon tone index are converted into return series, i.e.,$\textit{carbon price returns}_t = (\textit{Price of EUA}_t - \textit{Price of EUA}_{t-1})/\textit{Price of EUA}_{t-1}$ for $t = 2, 3, \cdots, T$ We also provide the descriptive statistics and data sources of dependent and independent variables in Table A.1 and draw the time series plots of all variables in Fig. A.1.

3.2. Carbon tone index

To calculate carbon tone index, our primary task is to create a keyword list. We carefully read 100 randomly selected news articles. Loughran and Mcdonald (2016) indicated that the word list should take the strategy focusing on unambiguous words rather than that containing many words broadly. We select all keywords that could affect changes in carbon prices, depending on marginal abatement cost theory and psychology. Table 1 reports all 198 keywords obtained from news articles. The first column lists 97 keywords that are expected to increase carbon prices. The second column reports 101 keywords that are expected to reduce carbon prices. The word list consists of keywords that we analysis and variants based on the root of keywords (e.g., boost, boosted, boosting, boosts). In the word list constructed in this article, we not only include keywords that reflect positive and negative sentiments based on psychology (e.g., hit, fell, rise, increase, fall), but also include keywords that affect the supply and demand of allowances (e.g., decarbonization, supplycurbing, brexit, delay, phaseout). Table 2 shows the fractional and cumulative percentages of the 30 most frequent keywords in all news. The top 30 keywords account for 84.82% of the cumulative count.

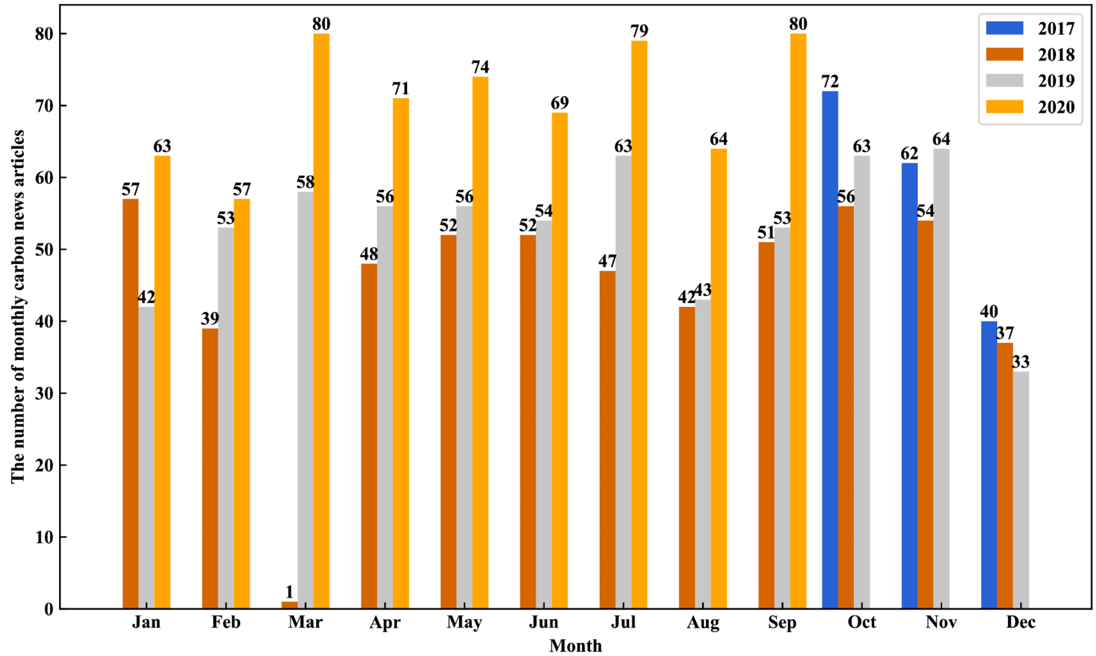

Fig. 1. The number of monthly carbon news articles.

Table 1
The list of 198 keywords created based on the vertis and carbon-pulse websites is expected to affect carbon price.

97 Keywords that could increase carbon prices				101 Keywords that could decrease carbon prices			
Above	Fine	Push	Spiked	Bearish	Delaying	Phaseout	Slipping
Approval	Fined	Pushed	Spikes	Below	Delays	Plummet	Slips
Approvals	Hit	Pushes	Spiking	Brexit	Dip	Plummeted	Slump
Approve	Hits	Pushing	Strengthen	Collapse	Dipped	Plummeting	Slumped
Approved	Hitting	Raise	Strengthened	Collapsed	Dipping	Plummets	Slumping
Approves	Increase	Raised	Strengthening	Collapses	Dips	Plunge	Slumps
Approving	Increased	Raises	Strengthens	Collapsing	Drop	Plunged	Steepen
Boost	Increases	Raising	Supplycurbing	Coronavirus	Dropped	Plunges	Steepened
Boosted	Increasing	Rebound	Support	Covid	Dropping	Plunging	Steepening
Boosting	Jump	Rebounded	Supported	Covid-19	Drops	Postpone	Steepens
Boosts	Jumped	Rebounding	Supporting	Curb	Exit	Postponed	Struggle
Bounce	Jumping	Rebounds	Supports	Curbed	Exited	Postpones	Struggled
Bounced	Jumps	Recover	Surge	Curbing	Exiting	Postponing	Struggles
Bounces	Lift	Recovered	Surged	Curbs	Exits	Reduce	Struggling
Bouncing	Lifted	Recovering	Surges	Cut	Fall	Reduces	Tumble
Bull	Lifting	Recovers	Surging	Cuts	Fallen	Reduced	Tumbled
Bulled	Lifts	Resurgence		Cutting	Falling	Reducing	Tumbles
Bulling	Nudge	Rise		Dampen	Falls	Reject	Tumbling
Bullish	Nudged	Risen		Dampened	Fell	Rejected	Virus
Bulls	Nudges	Rises		Dampening	Glut	Rejecting	Weakness
Buying	Nudging	Rising		Dampens	Gluts	Rejects	
Climb	Optimism	Rose		Decrease	Glutted	Sank	
Climbed	Optimistic	Soar		Decreased	Glutting	Sink	
Climbing	Peak	Soared		Decreases	Halt	Sinking	
Climbs	Peaked	Soaring		Decreasing	Halted	Sinks	
Decarbonisation	Peaking	Soars		Delay	Halting	Slip	
Double	Peaks	Spike		Delayed	Halts	Slipped	

The top five keywords that appear in the cumulative count table are above, support, below, hit, fell. The word support implies positive policy environment in EU ETS. In other words, it indicates the trend of carbon price, which directly induce changes in investor sentiment.

Additionally, we use the relative measure to calculate carbon tone index, which guarantees time series stationarity and avoids the case that all positive or negative words due to writing style (Kearney and Liu, 2014). Hence, carbon tone index for each article equals (the number of increasing words−the number of decreasing words)/(the number of words in the article). Attentionally, in order to extract useful content effectively, we implement text preprocessing before extracting keywords. We drop meaningless words such as single character and stopwords in text preprocessing. In addition, news more than one are published in most transaction days, so we obtain carbon tone index through the common daily average method. Soo (2018) extra verified the combined daily news and found that there is no significant difference between the two methods. In Appendix B, we list five news articles about carbon markets and point out the role of keywords in news.

3.3. Latent Dirichlet Allocation

Investors form different price expectations for different events.

Table 2
Most frequent keywords in the news articles from vertis and carbon-pulse websites.

	% of total key word count	Cumulative %		% of total key word count	Cumulative %
Above	7.36%	7.36%	Bullish	2.32%	64.81%
Support	7.13%	14.49%	Drop	2.30%	67.11%
Below	6.65%	21.14%	Recover	2.23%	69.34%
Hit	5.06%	26.20%	Climb	1.92%	71.26%
Fell	4.81%	31.01%	Buying	1.85%	73.11%
Rise	4.21%	35.22%	Coronavirus	1.75%	74.86%
Increase	4.18%	39.40%	Reduce	1.49%	76.35%
Fall	3.81%	43.22%	Dipped	1.40%	77.75%
Bearish	3.54%	46.76%	Jumped	1.39%	79.14%
Brexit	3.22%	49.98%	Raise	1.36%	80.50%
Push	2.73%	52.71%	Surge	0.92%	81.42%
Lift	2.57%	55.28%	Rebound	0.92%	82.33%
Cut	2.48%	57.76%	Double	0.89%	83.22%
Slip	2.37%	60.14%	Exit	0.83%	84.05%
Climbed	2.35%	62.49%	Decrease	0.77%	84.82%

Note: This table shows the fractional and cumulative percentages of the top 30 frequent keywords.

Therefore, we take Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to classify news articles and explore the influence of carbon tone index on EUA return under different news topics. Data are collected from August 5, 2011 to October 9, 2020.[3] We use the genism package (Hoffman et al., 2010) with Expectation-Maximum (EM) algorithm from python library for implementing LDA model. The basic process of LDA model is showed in Fig. 2. The target in LDA is to determine the topic distribution of each document and the distribution of words in each topic. Based on Bayesian theory, we assume the prior distribution of topics in any document $\theta_d = Dirichlet(\overrightarrow{\alpha})$ and the prior distribution of words in any topic $\beta_k = Dirichlet(\overrightarrow{\eta})$. α and η as hyperparameters in distribution are known. Correspondingly, the probability distribution of η word in d document from topic distribution θ_d is $z_{dn} = multi(\theta_d)$, and the probability distribution of word based on Z_{dn} is $w_{dn} = multi(\beta_{zdn})$. So, we obtain the posterior distribution of topic and word based on the Dirichlet distribution depend on Bayesian inference. N,D,K denotes the number of words, news articles and topics in corpus, respectively. We select the best parameter K according to the classified results. Importantly, after every news article is assigned a corresponding topic label, we calculate topic i carbon tone index on day t as Eq. (1), otherwise topic i carbon tone index on day t is equal to zero without any topic i news.

$$topic\ i\ carbon\ tone\ index_t = \frac{\sum carbon\ tone\ index\ of\ topic\ i\ on\ day\ t}{\sum news\ articles\ of\ topic\ i\ on\ day\ t} \tag{1}$$

3.4. Sample period division

We show the price and volume plots of EUA in Fig. 3. The full sample period covers the latest round of price increases since 2018. Some scholars indicate that changes in the expectation of carbon market are caused by the approved MSR that affects the number of surplus allowances in the EU ETS (Friedrich and Pahle, 2020). We draw a red straight line on January 1, 2019 in Fig. 3. Specially, MSR came into effect in January 2019 and absorbed about a quarter of the estimated aggregate allowances in circulation.[4] We clearly observe a continuously rising round of EUA price with high volume on the left side of the red line. The volume on the right side of the red line keeps very sluggish except the period from November 26, 2019 to December 16, 2019, during which the heads of government discussed and approved a non-binding 2050 climate neutrality plan at several European Council Meetings. Stricter abatement targets may lead to a surge in trading volume. We expect that the increased volume before the implementation of MSR will change the market's response speed to information. Hence, we design three sample periods: 296 samples from September 19, 2017 to January 1, 2019; 441 samples from January 1, 2019 to October 9, 2020; 737 samples from September 19, 2017 to October 9, 2020.

4. Empirical results and discussion

4.1. Topic

Relying on the visualization results of topic analysis, when the number of topics is equal to 5, these topics are clearly segmented and have the explicit practical meaning. Table 3 shows top 20 topic words with the largest weights in each topic in LDA model. The words in bold frankly indicate the realistic topic content. We hold that Topic 1 closely links with the auction market scheme and transaction environment based on the words below, auction and strong. Topic 2 may suggest the supply and demand level of carbon allowances because it includes the words allowance, benchmark and increase. Topic 3 involves many driving factors based on the words emission, coal and power. According to the words such as commission, reform and parliament, we speculate that Topic 4 reflects the policy-making about market reform and climate target agendas in EU ETS. The dma (Different of Moving Average) in topic 5 is a technical indicator and the other words price, above and signal also imply the technical analysis content. The realistic topic content will help us understand the estimation results in Section 4.2.

4.2. OLS regression

We use OLS regression to test the relationship between carbon price return and carbon-related news in EU ETS. The basic model specification is shown in Eq. (2), where we do not consider any control variables. The entity of keywords such as rise and fall may point to some control variables, which may lead to insignificant carbon tone index. $Lag(\cdot)$ denotes the lag operator. Particularly, lag zero order denotes $carbon\ tone\ index_t$. $Lag(carbon\ tone\ index)$ includes one or several $carbon\ tone\ index_{t-i}$, $i = 0, 1, 2, \cdots, 10$. β is the corresponding coefficient vector.

$$carbon\ price\ return_t = \alpha + \beta Lag(carbon\ tone\ index_t) + \varepsilon_t \tag{2}$$

Table 4 shows the estimation results of the lagged carbon tone variables on carbon price return in OLS regression. We only report the estimation results that have significance at the 10% level. We use standard errors with Newey-West correction for heteroscedasticity and autocorrelation with up to 5 lags. The robust statistic t is reported under the estimated coefficients. The number of observations is shown in the first column in Table 4. In the full sample regression, Table 4 shows that the coefficient of $carbon\ tone\ index_{t-8}$ is 0.0590 in column [1], which is statistically significant at the 5% level (t statistic is 2.1465). In column [2], we add the contemporary $carbon\ tone\ index_t$, and then the coefficient of $carbon\ tone\ index_{t-8}$ remains 0.0590, which is statistically significant at the 5% level (t statistic is 2.1443). We find that two carbon tone indexes both have a positive impact on carbon price return and generate the Adj. R^2 of 0.0047. The higher carbon tone index denotes the more positive sentiment in the news article. Hence, the estimation results seem reasonable. Some studies found that traders and investors have overreaction in stock and crude oil markets. In other words, asset return generally shows a reversal effect and the coefficient of sentiment index is negative at $t-1$. In our regression, carbon return shows positive reaction to news content at $t-8$, which appears to imply the low efficiency in EU ETS, that is, slow information absorption. Similarly, in the early stages of the stock market, U.S. stocks took a week to complete the reversal process of asset returns in the late 1990s (Tetlock, 2007). In the subsample period from September 19, 2017 to January 1, 2019, the results show that $carbon\ tone\ index_{t-4}$, $carbon\ tone\ index_{t-7}$ and *carbon*

[3] We use all available news data in LDA model to help categorize news topics.
[4] The report is collected from https://www.refinitiv.com/en/resources/special-report/review-of-carbon-markets-in-2018.

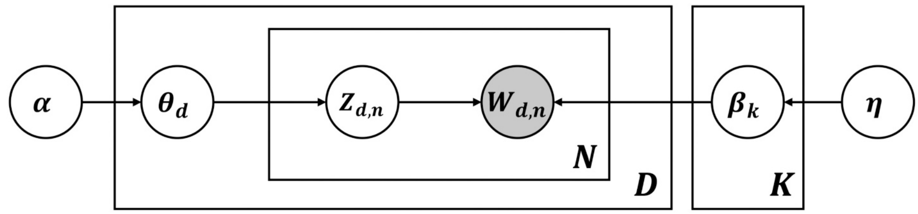

Fig. 2. Latent Dirichlet Allocation model.

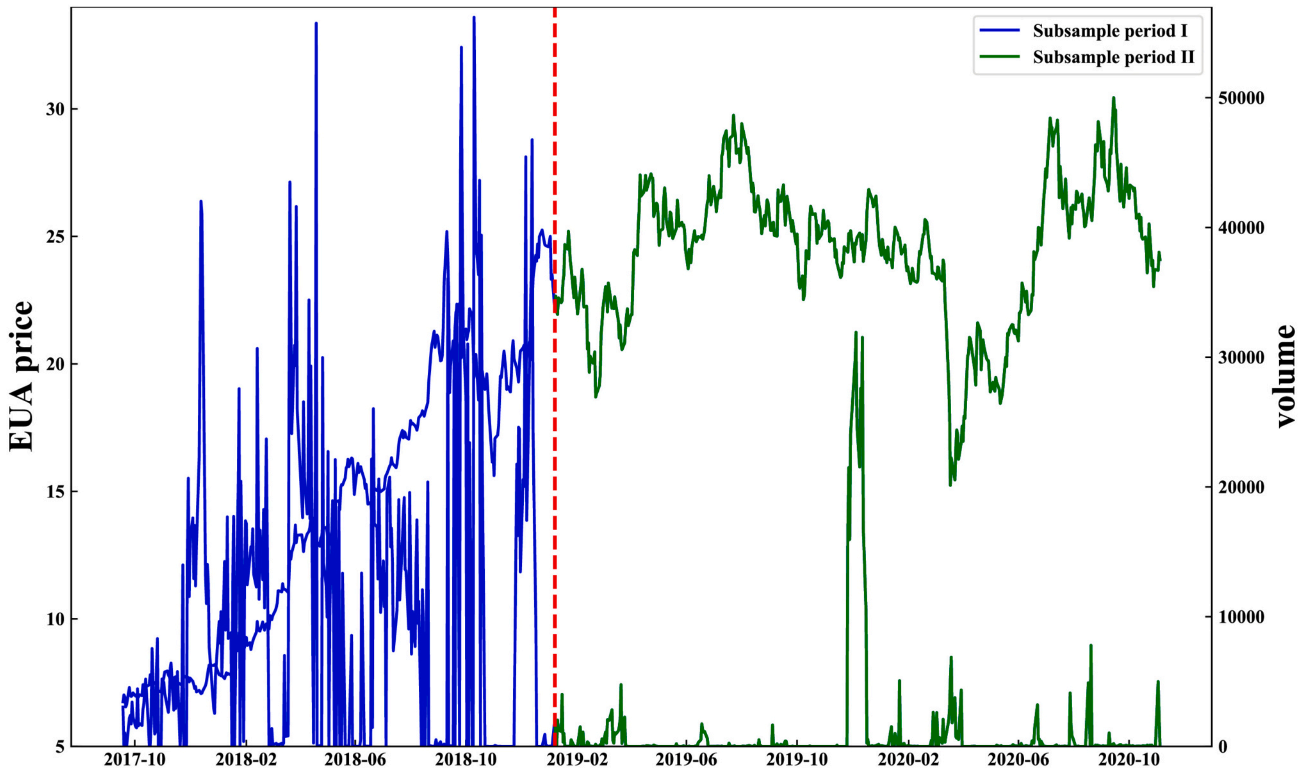

Fig. 3. The price and volume plots of EUA.

tone index$_{t-8}$ all have significant effects on carbon price return. Unexpectedly, *carbon tone index*$_{t-4}$ shows a negative 0.0584 (t statistic is -1.6639) at 10% significance level. We speculate that this result may be due to the misinterpretation of market information by some market participants. Similarly, *carbon tone index*$_t$ also shows a negative 0.0595 in column [4]. In column [5] and [6] including *carbon tone index*$_{t-8}$, only *carbon toneindex*$_{t-8}$ is statistically significant at the 5% level and still maintains a positive effect on carbon return, which may indicate that *carbon tone index*$_t$, *carbon tone index*$_{t-4}$ and *carbon tone index*$_{t-7}$ have weak influence. During the subsample period from January 1, 2019 to October 9, 2020, the results show an overreaction consistent with the empirical literature. In column [2], *carbon tone index*$_t$ is associated with an 0.0382 basis point higher return and *carbon tone index*$_{t-1}$ is associated with an 0.0634 basis point lower return. In addition, *carbon tone index*$_{t-8}$ has no significance during this period. We cautiously give a guess that carbon market became more efficient after a round of high trading volume. We will use quantile regression in Section 4.3 to further verify the robustness of the results. Compared with the 0.24% explanatory

Table 3
Topic words with the largest weights in each topic in LDA model.

Tag	Topic proportion	Top 20 topic words with the largest weights	The realistic topic content
Topic 1	9.07%	level, **below**, **auction**, contract, daily, but, end, start, **strong**, friday, **decline**, trade, carbon, **down**, first, european, energy, follow, get, trader	The auction market schemes and auction transaction environment
Topic 2	44.46%	be, market, have, carbon, not, **allowance**, new, eu, trading, year, **benchmark**, there, **increase**, **more**, but, also, **level**, break, time, range	The supply and demand level of allowance
Topic 3	7.01%	german, have, **emission**, go, bloomberg, profit, account, phase, **coal**, january, march, **power**, lose, installation, december, here, hand, base, bank, period	Fundamental factors
Topic 4	17.55%	ets, **commission**, eu, **reform**, emission, source, look, vote, European, **parliament**, meeting, **climate**, member, country, committee, **council**, state, **agreement**, **target**, political	Policy-making in EU ETS
Topic 5	21.91%	**price**, euro, day, low, high, week, **eua**, dec, **above**, last, close, support, cent, eur, **dma**, back, **signal**, **resistance**, **hit**, **fell**	Technical analysis

Note: The words in bold with the significant meanings point to the realistic topics.

Table 4
The table shows the estimation results in OLS regression with Newey-West standard errors: the slope coefficients of carbon tone index for the three sample periods of the EU ETS.

Phase	Independent variable	Dependent variable: carbon price return [1]	[2]	[3]	[4]	[5]	[6]
Full sample	Intercept	0.0020* (1.7223)	0.0020* (1.7022)				
2017/9/19–2020/10/	*carbon tone index$_t$*		0.0032 (0.1504)				
09	*carbon tone index$_{t-8}$*	0.0590** (2.1465)	0.0590** (2.1443)				
Observations:737	Adj. R^2	0.0060	0.0047				
Subsample	Intercept	0.0058*** (2.9401)	0.0037** (2.2047)	0.0037** (2.0539)	0.0066*** (3.2078)	0.0046** (2.4458)	0.0043** (2.3931)
2017/9/19–2019/01/	*carbon tone index$_t$*				−0.0595* (−1.6546)	−0.0564 (−1.5685)	−0.0504 (−1.3574)
01	*carbon tone index$_{t-4}$*	−0.0584* (−1.6639)			−0.0597* (−1.6909)		−0.0566 (−1.5702)
Observations:296	*carbon tone index$_{t-7}$*		0.0931* (1.6660)				0.0793 (1.3937)
	carbon tone index$_{t-8}$			0.0899** (2.4108)		0.0857** (2.2391)	0.0737** (2.0414)
	Adj. R^2	0.0025	0.0121	0.0110	0.0054	0.0134	0.0227
Subsample	Intercept	0.0004 (0.2875)	0.0005 (0.3322)				
2019/01/01–2020/	*carbon tone index$_t$*		0.0382 (1.3490)				
10/09	*carbon tone index$_{t-1}$*	−0.0596* (−1.7385)	−0.0634* (−1.8333)				
Observations:441	Adj. R^2	0.0063	0.0074				

Note: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1%level.

power of negative words in stock market (Tetlock et al., 2008) and the 0.34% explanatory power of oil tone index in oil market (Loughran et al., 2019), carbon tone index has a higher R^2.

Next, we discuss which topic content will affect carbon price return in Table 5. The basic model specification is shown in Eq. (3). Likewise, we do not consider any control variables and consider only single factor in the regression. In Section 4.1, we introduce the realistic content for different topic carbon tone indexes. During the full sample period from September 19, 2017 to October 9, 2020, *topic* 2 *carbon tone index$_{t-7}$* and *topic* 2 *carbon tone index$_{t-8}$* have positive 0.1596 and 0.1672 effects respectively on carbon price return, and have significance at the 1% level (t statistics are 2.7917 and 2.7065 respectively). Consistent with the findings summarized in (Hintermann et al., 2016), the information affecting the supply and demand of allowances plays a leading role in carbon price changes. In the subsample period from January 1, 2019 to October 9, 2020, *topic* 2 *carbon tone index$_{t-1}$* has a negative 0.1113 at the 10% level, which results from the fast overreaction of market participants like other financial markets. Compared with the subsample period from September 19, 2017 to January 1, 2019, *topic* 2 *carbon tone index$_{t-1}$* is an emerging driver for carbon pricing. Topic 1 generally summarizes the auction market situation. The positive market response will provide sufficient allowances supply for carbon market, which causes subsequent negative effects in the two subsample periods. The carbon market has an earlier negative reaction for *topic* 1 *carbon tone index* in the latter subsample period than that in the former subsample period. At first glance,*topic* 3 *carbon tone index$_{t-6}$* has statistically opposite significant coefficient at the 5% level in the two subsample periods. This result may be explained because news articles about fundamentals cover crude oil, natural gas and electricity, and there are some opposite effects among them. This finding indicates the future research direction that optimizing the dictionary to capture more detailed information. Some important issues, such as the more ambitious climate goals and the Brexit process, were discussed and determined during the period from January 1, 2019 to October 9, 2020. Therefore, we capture a positive 0.2807 coefficient of *topic* 4 *carbon tone index$_{t-1}$* on carbon price return at the 10% level. Carbon price return also showed faster negative reactions to new articles about technical analysis in the latter subsample period. Most of news articles about technical analysis are post-analysis, that is, the summary of historical market conditions and the analysis of changes in carbon price. Hence, the reversal effect always appears after the cumulative return of multiple days. The lagged *topic* 5 *carbon tone index* all produced a negative effect on carbon price return at the 10% level in two subsample periods. In general, the changes between the two subsample periods indicate faster information absorption in EU ETS. However, this may be also the reason that some topic carbon tone indexes have no significance in the full sample period. In short, the news media provides investors with interesting content and causes changes in carbon prices.

$$carbon\ price\ return_t = \alpha + \beta Lag(topic\ i\ carbon\ tone\ index) + \varepsilon_t\ \mathrm{i} = 1, 2, 3, 4, 5 \tag{3}$$

4.3. Quantile regression

Quantile regression, proposed by Koenker and Bassett Jr (1978), is mainly used to study the correlation between financial variables. We build the quantile regression of the conditional quantiles of dependent variable (carbon price return) to independent variables (carbon price drivers), and the coefficient represents the effect of the independent variables on the dependent variable at each quantile. We set nine quantiles (τ=0.1, 0.2, …, 0.9) and add mean regression model as benchmark according to the settings in Tan and Wang (2017). Our control variables include fundamental factors and lagged carbon returns based on partial autocorrelation plot in quantile regression. We choose carbon tone index that has robust significance at the 10% level in Section 4.2. The estimation results in Table 6 show that carbon tone indexes

Table 5
The table shows the estimation results in OLS regression with Newey-West standard errors: the slope coefficients of topic carbon tone index for the three sample periods of the EU ETS.

Phase	Independent variable	Dependent variable: carbon price return [1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Full sample 2017/9/19–2020/10/09 Observations:737	Intercept	0.0020* (1.7821)	0.0020* (1.7286)	0.0018 (1.5523)					
	topic 2 *carbon tone index*$_{t-7}$	0.1596*** (2.7917)		0.1447*** (2.6134)					
	topic 2 *carbon tone index*$_{t-8}$		0.1672*** (2.7065)	0.1533** (2.5785)					
	Adj. R^2	0.0133	0.0148	0.0254					
Subsample 2017/9/19–2019/1/01 Observations:296	Intercept	0.0051*** (2.6278)	0.0035** (2.1604)	0.0036** (1.9783)	0.0046** (2.3757)	0.0061*** (3.2020)	0.0055*** (2.8455)	0.0058*** (2.9818)	
	topic 1 *carbon tone index*$_{t-9}$	−0.2067** (−2.4555)							
	topic 2 *carbon tone index*$_{t-7}$		0.2827* (1.9235)						
	topic 2 *carbon tone index*$_{t-8}$			0.2787*** (2.8139)					
	topic 3 *carbon tone index*$_{t-6}$				−0.4082** (−2.1489)				
	topic 5 *carbon tone index*$_{t}$					−0.1375** (−2.2653)			
	topic 5 *carbon tone index*$_{t-2}$						−0.0800* (−1.7472)		
	topic 5 *carbon tone index*$_{t-4}$							−0.0817* (−1.7959)	
	Adj. R^2	0.0062	0.0292	0.0283	0.0071	0.0160	0.0031	0.0033	
Subsample 2019/1/01–2020/10/09 Observations:441	Intercept	0.0003 (0.2105)	0.0003 (0.2297)	0.0005 (0.3533)	0.0006 (0.4167)	0.0009 (0.6248)	0.0006 (0.4279)	0.0004 (0.2730)	0.0006 (0.4315)
	topic 1 *carbon tone index*$_{t-4}$	−0.1488* (−1.7243)							
	topic 1 *carbon tone index*$_{t-7}$		−0.1335* (−1.6942)						
	topic 2 *carbon tone index*$_{t-1}$			−0.1113* (−1.8753)					
	topic 2 *carbon tone index*$_{t-7}$				0.0952** (2.1152)				
	topic 3 *carbon tone index*$_{t-6}$					0.4395** (2.1917)			
	topic 4 *carbon tone index*$_{t-1}$						0.2807* (1.7152)		
	topic 5 *carbon tone index*$_{t}$							0.0923* (1.7939)	
	topic 5 *carbon tone index*$_{t-1}$								−0.0851* (−1.8549)
	Adj. R^2	0.0060	0.0044	0.0063	0.0040	0.0081	0.0043	0.0059	0.0060

Notes: Topic 1 carbon tone index covers the auction market schemes and transaction environment content. Topic 2 carbon tone index covers the supply and demand level of allowance content. Topic 3 carbon tone index covers the fundamental factors content. Topic 4 carbon tone index covers the policy-making content in EU ETS. Topic 5 carbon tone index covers the technical analysis content. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level.

all have no asymmetric effect and maintains the same sign at any quantile. Interestingly, carbon tone index usually has a statistically significant and higher effect at the tail quantiles. Consistent with the rapid overreaction in the subsample period from January 1, 2019 to October 9, 2020 in Section 3.2, carbon tone index has a negative 0.0641 impact on carbon prices at the 5% significance level at the 0.1 quantile and has a negative 0.1464 negative impact on carbon prices at 1% significance level at the 0.9 quantile. Common control variables also show significant effects on carbon price return in different periods. Electricity, coal and crude oil prices have significant influences on carbon prices at the 10% level in the former subsample period. Carbon prices did not exceed the fuel switch price[5] (Koch et al., 2014) during most of this period and therefore are mainly affected by the energy consumption (emissions) that electricity and crude oil prices reflected. Differently, stock and gas prices show significant effects on carbon prices at the 10% level in the latter subsample period. Carbon prices exceeded the fuel switch price during this period and hence natural gas price was verified as an important factor. In addition, the Covid-19 that broke out in 2019 severely hinder economic growth, so stock market, which is regarded as an economic barometer, have a significant effect on carbon prices. To sum up, the estimation results keep consistency with that in Section 4.2. We also find that carbon tone index has statistically significant effects on carbon price return at low and high tail quantiles.

[5] The switching price denotes the abatement opportunities through fuel switching by calculating the allowance price at which the marginal costs of gas and coal power plants are equal.

Table 6
The table shows the estimation results in OLS and Quantile regression with Newey-West standard errors: the slope coefficients of all independent variables for the three sample periods of the EU ETS.

Phase	Independent Variable	OLS	Dependent variable: carbon price return								
			$Q_{0.1}$	$Q_{0.2}$	$Q_{0.3}$	$Q_{0.4}$	$Q_{0.5}$	$Q_{0.6}$	$Q_{0.7}$	$Q_{0.8}$	$Q_{0.9}$
Full sample	Intercept	0.0020	−0.0317***	−0.0189***	−0.0111***	−0.0042***	0.0016	0.0078***	0.0141***	0.0243***	0.0371***
2017/9/19–2020/		(1.6386)	(−13.2779)	(−12.990)	(−9.1091)	(−3.4828)	(1.3850)	(6.7365)	(11.0430)	(15.2539)	(14.2247)
10/09	*carbon tone*	0.0620**	0.0928**	0.0497*	0.0189	0.0350	0.0590**	0.0651**	0.0513	0.0363	0.0519
Observations:737	$index_{t-8}$	(2.1987)	(2.1679)	(1.7054)	(0.8197)	(1.4845)	(2.3042)	(2.2368)	(1.4381)	(1.0918)	(0.8512)
	$Electricity_{t-1}$	0.0367	0.2138 (1.2127)	−0.0196	−0.0307	−0.1259	−0.1683	−0.1749*	−0.1653	0.0259	−0.0224
		(0.3345)		(−0.1754)	(−0.3030)	(−1.1413)	(−1.6356)	(−1.8266)	(−1.4553)	(0.1831)	(−0.1138)
	$Oil_{t\text{-}1}$	0.0229	0.0837 (1.5500)	0.0236 (0.5945)	−0.0238	0.0064	−0.0071	−0.0603	−0.0656*	−0.0292	0.0544
		(0.5017)			(−0.5328)	(0.1615)	(−0.1622)	(−1.3611)	(−1.7109)	(−0.5228)	(1.4568)
	$Coal_{t-1}$	−0.0718	−0.2051***	−0.1229***	−0.0445	−0.0016	−0.0182	0.0502	0.0587	0.0629	0.0425
		(−0.9199)	(−2.7682)	(−2.6430)	(−0.6625)	(−0.0226)	(−0.2685)	(0.7538)	(0.9722)	(0.9582)	(0.5878)
	$Stoxx_{t-1}$	−0.2089	−0.1275	0.0542 (0.7152)	0.0425	−0.0069	−0.0749	−0.0302	−0.0128	−0.2936	−0.4617*
		(−1.5939)	(−1.2891)		(0.3427)	(−0.0441)	(−0.3764)	(−0.1577)	(−0.0711)	(−1.2054)	(−1.8725)
	Gas_{t-1}	−0.0600*	−0.0695	−0.0103	−0.0263	−0.0392**	−0.0498**	−0.0498**	−0.0486	−0.0980***	−0.1116***
		(−1.8069)	(−0.8129)	(−0.5026)	(−1.4131)	(−1.9980)	(−2.4637)	(−2.2233)	(−1.2182)	(−5.1265)	(−6.1140)
	Adj. R^2	0.0131	0.0047	0.0010	−0.0024	0.0014	0.0086	0.0111	0.0108	0.0069	0.0064
Subsample	Intercept	0.0036**	−0.0312***	−0.0172***	−0.0084***	−0.0017	0.0048**	0.0101***	0.0151***	0.0239***	0.0336***
2017/9/19–2019/		(2.1092)	(−9.5774)	(−6.0393)	(−3.6672)	(−0.8459)	(2.5411)	(5.3947)	(8.0391)	(10.9262)	(13.0132)
1/01	EUA_{t-1}	0.2950***	0.3542***	0.2860***	0.2598**	0.1886**	0.2010**	0.2136***	0.2423***	0.2074***	0.1412
Observations:296		(3.8153)	(3.8438)	(2.8175)	(2.0392)	(2.3524)	(2.1779)	(2.9662)	(3.5984)	(3.3847)	(1.4753)
	carbon tone	0.0682*	0.0780*	0.0407 (1.1232)	0.0228	0.0485	0.0442	0.0885	0.1019*	0.1212**	0.2393***
	$index_{t-8}$	(1.7935)	(1.6924)		(0.6464)	(1.1670)	(0.8955)	(1.6475)	(1.7077)	(2.0829)	(3.5300)
	$Electricity_{t-1}$	−0.5020***	−0.5071***	−0.4039***	−0.5522**	−0.4324**	−0.4518**	−0.5656***	−0.7162***	−0.4582	−0.4882*
		(−2.8737)	(−3.5508)	(−2.7610)	(−2.4928)	(−2.2066)	(−2.2924)	(−3.0222)	(−4.0046)	(−1.5054)	(−1.7080)
	$Oil_{t\text{-}1}$	−0.1709*	−0.2104*	−0.0452	−0.0356	−0.0345	−0.0784	0.0083	−0.0449	−0.1718	−0.2026
		(−1.9495)	(−1.8192)	(−0.4421)	(−0.2498)	(−0.2895)	(−0.6297)	(0.0656)	(−0.3453)	(−0.9364)	(−1.2263)
	$Coal_{t-1}$	−0.3272**	−0.1267	−0.1824	−0.2670	−0.4000***	−0.3326**	−0.2457	−0.2623*	−0.5167**	−0.7126*
		(−2.4020)	(−0.8675)	(−1.2808)	(−1.3588)	(−2.9169)	(−2.1328)	(−1.5263)	(−1.7585)	(−2.1633)	(−1.9960)
	$Stoxx_{t-1}$	0.2711	0.4364 (1.1694)	0.2243 (0.5106)	0.3509	−0.0373	0.2054	0.0947	0.2074	0.4183*	0.3140
		(1.3742)			(0.7680)	(−0.1384)	(0.7393)	(0.3915)	(0.9282)	(1.7639)	(0.8879)
	Gas_{t-1}	−0.1346	0.1474*	−0.0734	−0.0936	−0.0393	−0.0845	−0.0707	−0.1072	−0.1612*	−0.0784
		(−0.8551)	(1.7526)	(−0.6125)	(−0.7270)	(−0.4188)	(−0.9240)	(−0.8168)	(−1.3436)	(−1.8631)	(−0.4958)
	Adj. R^2	0.1000	0.0578	0.0307	0.0258	0.0248	0.0372	0.0523	0.0547	0.0685	0.0931
Subsample	Intercept	0.0005	−0.0349***	−0.0211***	−0.0124***	−0.0061***	−7.49E-05	0.0068***	0.0137***	0.0234***	0.0355***
2019/1/01–2020/		(0.3794)	(−13.6264)	(−10.6363)	(−7.3118)	(−3.7161)	(−0.0460)	(4.2307)	(7.4472)	(10.9292)	(15.1912)
10/09	EUA_{t-2}	0.1031	0.0784 (1.0839)	0.0601 (0.9548)	0.0370	0.0541	0.0740	0.0541	0.0406	0.0216	0.0147
Observations:441		(1.3889)			(0.5949)	(0.8161)	(1.2168)	(0.9336)	(0.6578)	(0.3892)	(0.2272)
	EUA_{t-6}	−0.1010*	−0.1109*	−0.0298	−0.0292	−0.0747	−0.1144*	−0.1056	−0.0946	−0.0917	−0.1676**
		(−1.7961)	(−1.8259)	(−0.6499)	(−0.5787)	(−1.1033)	(−1.7513)	(−1.5235)	(−1.2852)	(−1.1176)	(−2.0187)
	carbon tone	−0.0547	−0.0641**	−0.0599	−0.0231	−0.0299	−0.0108	−0.0284	−0.0661	−0.0529	−0.1464***
	$index_{t-1}$	(−1.6450)	(−1.9735)	(−1.1966)	(−0.5677)	(−0.8080)	(−0.2891)	(−0.7347)	(−1.5146)	(−1.2281)	(−3.0800)
	$Electricity_{t-1}$	0.0999	0.4507**	0.2589*	0.0445	0.0279	−0.0225	−0.0638	−0.0141	0.0654	0.1264
		(0.8211)	(2.0637)	(1.6875)	(0.2991)	(0.1900)	(−0.1585)	(−0.4911)	(−0.1164)	(0.5294)	(0.8731)
	$Oil_{t\text{-}1}$	0.0359	0.1202**	0.0505 (1.0427)	−0.0162	−0.0359	−0.0549	−0.1088***	−0.0768*	0.0063	0.1071**
		(0.7425)	(2.2359)		(−0.3068)	(−0.6713)	(−1.0000)	(−2.7381)	(−1.6852)	(0.1064)	(2.4227)
	$Coal_{t-1}$	0.0026	−0.2672***	−0.1578**	−0.0376	0.0201	0.0462	0.1432*	0.1426**	0.0904**	0.0670
		(0.0316)	(−2.8714)	(−2.4402)	(−0.3617)	(0.1532)	(0.3990)	(1.7788)	(2.0360)	(2.0362)	(1.5165)
	$Stoxx_{t-1}$	−0.2742*	−0.1644*	−0.0246	0.0003	0.0334	0.0068	0.0854	−0.0424	−0.3202	−0.4556**
		(−1.7593)	(−1.7380)	(−0.2679)	(0.0019)	(0.1931)	(0.0326)	(0.4342)	(−0.1859)	(−1.3107)	(−2.1575)
	Gas_{t-1}	−0.0462*	−0.0785	−0.0032	−0.0178	−0.0421*	−0.0475	−0.0419	−0.0589**	−0.0813***	−0.0750***
		(−1.7179)	(−0.7053)	(−0.1289)	(−0.7797)	(−1.7095)	(−1.9607)	(−1.6363)	(−2.4147)	(−3.4539)	(−4.0340)
	Adj. R^2	0.0281	0.0352	−0.0016	−0.0107	−0.0100	−0.0066	0.0030	0.0093	0.0212	0.0515

Note: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1%level.

Table 7
The prediction metric results for eight predictive models.

	Metrics	OLS_1	OLS_2	PCR	PLS	ElasticNet	SVR	ANN	GBRT
Fundamentals and sentiment	R^2	0.00137	−0.00641	**0.03573**	0.00449	0.00746	0.00991	−0.00440	0.01207
	MSE	0.00106	0.00107	**0.00103**	0.00106	0.00106	0.00105	0.00107	0.00105
	DA	0.45945	0.50000	0.54054	0.52703	0.44595	0.45946	**0.60811**	0.52703
Fundamentals	R^2	−0.59226	−0.51171	0.01647	−0.05494	−0.00348	0.00948	−0.01957	−0.00379
	MSE	0.00170	0.00161	0.00105	0.00112	0.00107	0.00106	0.00109	0.00107
	DA	0.58108	0.55405	0.51351	0.39189	0.40541	0.47297	0.47297	0.55405

Notes: OLS_1 only includes carbon tone indexes in quantile regression. OLS_2 includes all carbon tone indexes and topic carbon tone indexes that have significant effect on carbon price. The other models include all carbon tone index and topic carbon tone index that have significant effect on carbon price. The values in bold denote the best metrics in all models.

4.4. *A simple trading strategy based on all drivers*

The above empirical findings are applied to carbon price prediction in this paper. All significant carbon tone indexes and topic carbon tone indexes are regarded as carbon price predictors at the 10% level. We use classic machine learning models with reference to Gu et al. (2020): Principal Components Regression (PCR), Partial Least Squares (PLS), ElasticNet, Support Vector Regression (SVR), Artificial Neural Network (ANN) and Gradient Boost Regression Tree (GBRT). Considering many predictors, we include OLS_1 with only carbon tone indexes included in quantile regression and OLS_2 with all carbon tone indexes and topic carbon tone indexes. The former avoids the problems of multi-collinearity and reduced freedom degrees, which may increase estimation errors. Table 7 shows three metrics (the formulas of R^2, MSE and DA in Appendix C) in the two groups of predictors (Fundamentals and sentiment; Fundamentals) in all models. Fundamentals indicate all control variables (such as electricity and coal prices) used in Section 4.3. We show the predictors used in the predictive model in Table C.2and the division information of data set in Table C.3 in Appendix C. We find all models with sentiment indexes have better performance than that with only fundamentals. The dimension reduction method PCR, which eliminates noise through linear combination, obtains the best predictive result.

We evaluate economic value with the predictive results of PCR in market timing strategy. Considering the transaction cost 0.006 (Aatola et al., 2014) as threshold value, we buy (or sell) allowances at the close of t day and sell (or buy) it at the close of the $t+1$ day under the condition of $\left|\widehat{r}_{t+1}\right|$ is greater than 0.006, or we do nothing. The market timing strategy has the highest average annual return 47.69% and success ratio 63.16% in Table 8. In addition, it also has medium risk with standard deviation 0.47 and *MaxDD* 5.15%. Intuitively, we plot the cumulative return based on the assumption that the initial asset is 1. Fig. 4. shows that our strategy maintains a steady rising return.

Table 8
The economic value of market timing strategy.

Trading strategies	Average return (%)	Standard deviation	Success (%)	Sharpe ratio	*MaxDD* (%)
Market time	47.69%	0.47	63.16%	1.02	5.15%
Buy and hold	12.46%	0.52	50.00%	0.24	19.18%
Always long	18.27%	0.44	61.54%	0.42	3.73%

Notes: This table reports the economic value of the market timing strategy, compared with the benchmark buy and hold strategy and always-long strategy. Average return is annualized by multiplication by 252. Average return, success and *MaxDD* are expressed as a percentage.

5. Conclusions and policy implications

In this paper, we empirically find that carbon tone index based on the self-built dictionary affects carbon price return.*carbon tone index*$_{t-8}$ is statistically significant during the subsample period from September 19, 2017 to January 1, 2019, but *carbon tone index*$_{t-1}$ shows the over-reaction during the subsample period from January 1, 2019 to October 9, 2020. The dissemination of media news articles into carbon prices seems to become faster. The findings provide government decision makers with suggestions that using flexible structural instruments such as MSR and encouraging more financial investors to enter carbon market for improving market efficiency. The LDA model divides media news articles into five new topics. Topic carbon tone indexes show different influence patterns on carbon prices. The findings can help government respond to the expected market changes in advance and ensure the stable operation of carbon market.

In addition, carbon tone index has been empirically tested as carbon pricing driver. Therefore, investors can use carbon pricing drivers to construct investment strategies. We design a simple trading strategy to verify the economic value of carbon tone index. In general, our research helps stakeholders understand carbon price formation mechanism. There are still some shortcomings in our research. We don't explore the influence mechanism of sentiment on carbon prices in depth. Some discussions as follows: (1) the short-term carbon price changes driven by the sentiment of noise traders; (2) the long-term carbon price level changes triggered by market incremental information; (3) both have. In addition, the dictionary could be further improved. It may help us separate out pure information with identified entities (such as sentiment indicators from only MD&A text or IPO text) to explore the above discussions.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in *Energy Economics Journal*.

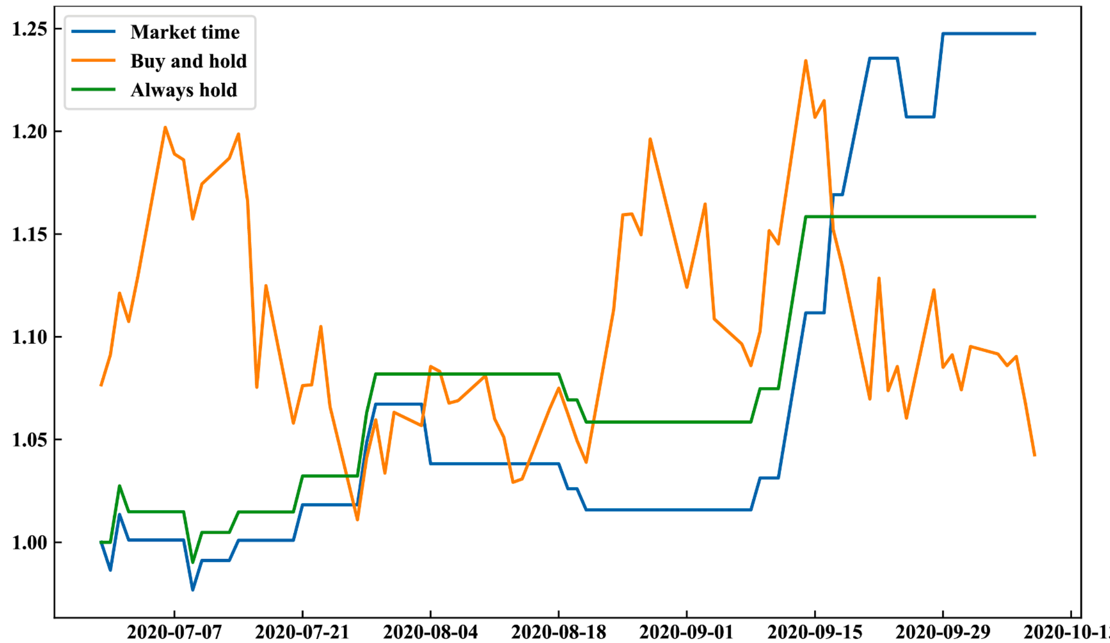

Fig. 4. The cumulative return of three trade strategies.

Appendix A. Data

Table A.1
Data statistics and correlations.

Panel A: Descriptive statistics

Acronym	Characteristic decription	Data source	Min	Max	Median	Mean	Std	ADF test	PP test
EUA	European Union Allowance	https://www.investing.com/	−0.1767	0.24802	0.0018	0.0023	0.0317	−26.56***	−26.60***
Tone	Carbon Tone Index	https://vertis.com/ https://carbon-pulse.com/	−0.1333	0.2308	0.0000	0.0041	0.0459	−24.36***	−24.76***
Gas	UK Natural Gas Future Price	https://www.quandl.com/	−0.1621	0.4336	−0.0003	0.0006	0.0423	−26.54***	−26.54***
Electricity	EEX Strom Phelix Baseload Year Future	https://markets.businessinsider.com/	−0.0740	0.0584	0.0003	0.0004	0.0143	−25.29***	−25.24***
Coal	Rotterdam Coal Future	https://www.investing.com/	−0.1651	0.1612	−0.0003	−0.0005	0.0168	−24.73***	−24.75***
Stoxx	STOXX Europe 600 Index	https://www.wsj.com/	−0.1148	0.0840	0.0008	0.0000	0.0117	−9.36***	−27.87***
Oil	Brent Oil Future Contract	https://www.investing.com/	−0.2519	0.2219	0.0009	0.0002	0.0302	−24.84***	−24.87***
ERIX	European Renewable Energy Total Return Index	https://www.investing.com/	−0.1217	0.0668	0.0018	0.0012	0.0154	−27.32***	−27.41***

Panel B: Correlation table of the variables

	EUA	Tone	Gas	Electricity	Coal	Stoxx	Oil
Tone	−0.0161						
Gas	−0.0906	0.0401					
Electricity	−0.0610	0.0048	0.4483				
Coal	−0.0647	0.0226	0.2645	0.2506			
Stoxx	−0.0453	−0.0684	0.0740	0.1898	0.1700		
Oil	−0.0404	−0.0780	0.1702	0.2183	0.2084	0.3107	
ERIX	0.0327	−0.0502	0.0342	0.1792	0.0644	**0.6096**	0.2334

Note: The collected data is from September 19, 2017 to October 9, 2020. The ADF and PP test denote the Dickey and Fuller (1979) and Phillips and Perron (1988) unit root tests separately. Value in bold denotes the strong correlation coefficient between two variables.

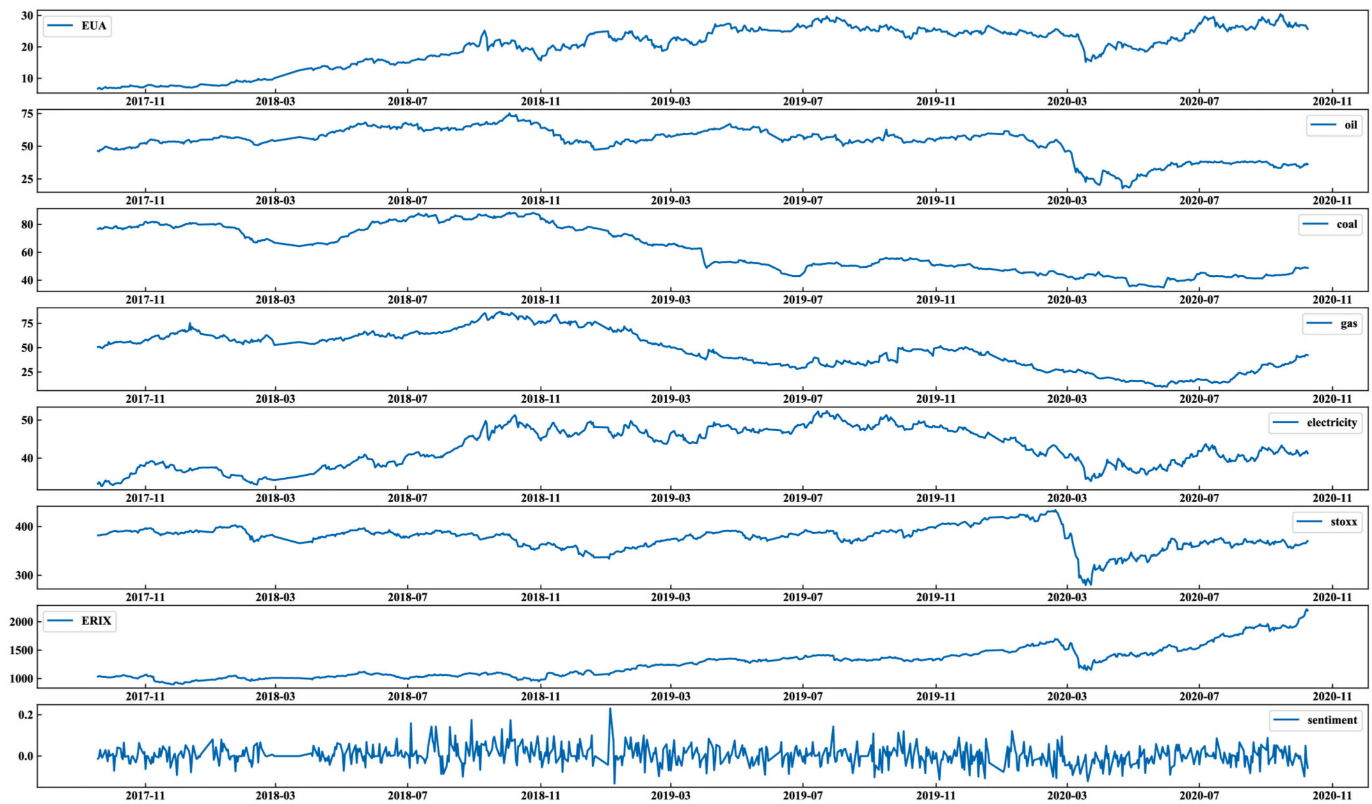

Fig. A.1. The time series plots for all variables

Appendix B. Some examples from the carbon-related new articles

In this section, we present five examples of carbon-related news articles. Words in bold indicate a decrease in carbon price while italicized words indicate an increase in carbon price.

1. Negative Example (Oct 09, 2020). Number of positive instances: 0 and number of negative instances: 2. Carbon tone index is −7.69% = (0–2)/26 words.
   UK-based EU ETS trading accounts will be **cut** off from the end of this year due to **Brexit** the British government confirmed this week while any holders of Kyoto Protocol credits will lose access to those units for several months.
2. Positive Example (Dec 13, 2019). Number of positive instances: 2 and number of negative instances: 1. Carbon tone index is 5.26% = (2–1)/19 words.
   EUAs on Friday ***lifted above*** the €25 handle that has drawn the market's focus this week with prices notching a 26% weekly gain despite some mounting **bearish** fundamental signals.
3. Negative Example (July 21, 2020). Number of positive instances: 2 and number of negative instances: 4. Carbon tone index is −9.52% = (2–4)/21 words.
   EUAs **sank** back **below** €24 on Wednesday though prices ended ***above*** technical ***support*** that some see as helping carbon defy wider **weakness** across financial markets on the impact of the **coronavirus.**
4. Positive Example (Jan 18, 2019). Number of positive instances: 2 and number of negative instances: 4. Carbon tone index is 9.09% = (2–0)/22 words.
   European carbon prices ***climbed*** to a fresh 2-month high on Thursday as ***bulls*** felt empowered after yesterday's options expiry saw a huge rally rather than the sell-off that some had foretold.
5. Positive Example (May 10, 2018). Number of positive instances: 2 and number of negative instances: 4. Carbon tone index is 4.55% = (0–1)/22 words.
   North American carbon markets endured an extremely quiet week with normal business in California interrupted by an aborted conference and the RGGI market **falling** silent due to an apparent lack of interest.

Appendix C. Three performance metrics, hyperparameters in the predictive models, the predictors in predictive models and the division information of data set

Three performance metrics

1. Mean square error (MSE)

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(r_t - \hat{r}_t\right)^2 \tag{C.1}$$

2. R^2

$$R^2 = 1 - \frac{\sum_{t=1}^{N}\left(r_t - \hat{r}_t\right)^2}{\sum_{t=1}^{N} r_t^2} \tag{C.2}$$

3. Directional accuracy (DA)

$$p_t = \begin{cases} 1, r_t{*}\hat{r}_t \geq 0 \\ 0, r_t{*}\hat{r}_t < 0 \end{cases} \tag{C.3}$$

$$DA = \frac{1}{N}\sum_{t=1}^{N} p_t \tag{C.4}$$

Table C.1
Hyperparameters for all models.

Model	Symbol	Meaning	Value	Range
PLS	K	The number of linear combinations	[0,1]	The number of predictors
PCR	K	The number of linear combinations	[14,2]	The number of predictors
ElasticNet	α	$l_1$norm penalty term	[0.01, 0.01]	[0.0001,0.001,0.01,0.1,0.5,1]
	λ	$l_2$norm penalty term	[0.001, 0.001]	[0.0001,0.001,0.01,0.1,0.5,1]
GBRT	n	The number of trees	[10,10]	[10,11, …,39,40]
	depth	The max depth of one tree	[7,5]	[5,6,7,8]
	split	The min samples split of one tree	[7,7]	[7,8, …,13,14]
	leaf	The min samples leaf of one tree	[25,40]	[20,21, …,69,70]
	r_1	Learning rate	[0.1, 0.1]	[0.01,0.05,0.1]
	subsample	The subsample ratio of one tree	[0.8,0.6]	[0.6,0.65,0.7,0.75,0.8,0.85,0.9]
ANN	A	Activation function	relu	relu
	bs	Size of batch	[8,8]	[8,16,32]
	h	Number of neurons	[14,21]	[10,11, …,49,50]
	α_2	$l_2$norm penalty term	[0.1,0.1]	[0.001,0.01,0.1]
SVR	γ	Radial basis function hyperparameter	[0.4,0.1]	[0.1, 0.2, 0.3,0.4,0.5,0.6,0.7,0.8,0.9,1]
	c	Penalty hyperparameter	[0.1,0.1]	[0.001, 0.01, 0.1, 1, 5, 10]

Note: the first hyperparameter is used for the fundamentals and sentiment group in the Value column; the second hyperparameter is used for the fundamentals group in the Value column.

Table C.2
The predictors in predictive models.

	Model	OLS_1	OLS_2/PCR/PLS/ElasticNet/SVR/ANN/GBRT
Predictor	Fundamentals	brent_return_lag_1,UKgas_return_lag_1,Coal_return_lag_1, Baseload_return_lag_1,stoxx_europe_600_return_lag_1	EUA_return_lag_1,EUA_return_lag_2,EUA_return_lag_6,brent_return_lag_1, UKgas_return_lag_1,Coal_return_lag_1,Baseload_return_lag_1, stoxx_europe_600_return_lag_1
	Fundamentals and sentiment	sentiment_lag_8,brent_return_lag_1,UKgas_return_lag_1, Coal_return_lag_1,Baseload_return_lag_1, stoxx_europe_600_return_lag_1	EUA_return_lag_1,EUA_return_lag_2,EUA_return_lag_6,sentiment_lag_1, sentiment_lag_4,sentiment_lag_7,sentiment_lag_8,topic_1_lag_4,topic_1_lag_7, topic_1_lag_9,topic_2_lag_1,topic_2_lag_7,topic_2_lag_8,topic_3_lag_6, topic_4_lag_1,topic_5_lag_1,topic_5_lag_2, topic_5_lag_4,brent_return_lag_1, UKgas_return_lag_1,Coal_return_lag_1, Baseload_return_lag_1, stoxx_europe_600_return_lag_1

Table C.3
The division information of data set.

Model	Training set	Validation set	Test set
OLS_1/OLS_2	655	–	74
PCR/PLS	581	74	74
	Train and validation set		
ElasticNet/SVR/ANN/GBRT	655 (three-fold cross-validation)		74

Note: There are no hyperparameters in OLS_1 and OLS_2 models so they have more training set and no validation set. Due to the limitation of small datasets, we divide training, validation and test sets according to 8/1/1 ratio in PCR and PLS models (This division allows as much data as possible to be used for training so that we can obtain a good prediction model.). We use grid search to determine the hyperparameters of the four models (ElasticNet/SVR/ANN/GBRT). The validation set is determined by the default setting (three-fold cross-validation) of the grid search function.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2021.105393.

References

Aatola, P., Ollikainen, M., Toppinen, A., 2013. Price determination in the EU ETS market: theory and econometric analysis with market fundamentals. Energy Econ. 36, 380–395.

Aatola, P., Ollikka, K., Ollikainen, M., 2014. Informational efficiency of the EU ETS market – a study of price predictability and profitable trading. J. Environ. Econ. Policy 3 (1), 92–123.

Alberola, E., Chevallier, J., Chèze, Bt, 2008. Price drivers and structural breaks in European carbon prices 2005–2007. Energy Policy 36 (2), 787–797.

Andrei, D., Hasler, M., 2014. Investor attention and stock market volatility. Rev. Financ. Stud. 28 (1), 33–72.

Association IET, 2012. GHG Market Sentiment Survey. PwC and IETA.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. J. Financ. Econ. 49 (3), 307–343.

Black, F., 1986. Noise. J. Financ. 41 (3), 528–543.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Brandt, M.W., Gao, L., 2019. Macro fundamentals or geopolitical events? A textual analysis of news events for crude oil. J. Empir. Financ. 51, 64–94.

Caporin, M., Poli, F., 2017. Building news measures from textual data and an application to volatility forecasting. Econometrics 5 (3), 35.

Chevallier, J., 2009. Carbon futures and macroeconomic risk factors: a view from the EU ETS. Energy Econ. 31 (4), 614–625.

Chevallier, J., 2011. Macroeconomics, finance, commodities: interactions with carbon markets in a data-rich model. Econ. Model. 28 (1), 557–567.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under-and overreactions. J. Financ. 53 (6), 1839–1885.

Deeney, P., Cummins, M., Dowling, M., Smeaton, A.F., 2016. Influences from the European Parliament on EU emissions prices. Energy Policy 88, 561–572.

Deveikyte, J., Geman, H., Piccari, C., Provetti, A., 2020. A sentiment analysis approach to the prediction of market volatility. arXiv preprint arXiv:2012.05906.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74 (366a), 427–431.

Duffie, D., 2010a. Asset price dynamics with slow-moving capital. J. Financ. 65, 1237–1267.

Duffie, D., 2010b. Presidential address: asset price dynamics with slow-moving capital. J. Financ. 65 (4), 1237–1267.

Dutta, A., Bouri, E., Noor, M.H., 2018. Return and volatility linkages between CO2 emission and clean energy stock prices. Energy 164, 803–810.

Fan, J.H., Todorova, N., 2017. Dynamics of China's carbon prices in the pilot trading phase. Appl. Energy 208, 1452–1467.

Ferris, S.P., Hao, Q., Liao, M.-Y., 2012. The effect of issuer conservatism on IPO pricing and performance. Rev. Financ. 17 (3), 993–1027.

Friedrich, M., Pahle, M., 2020. Allowance Prices in the EU ETS–Fundamental Price Drivers and the Recent Upward Trend. Working Paper. Potsdam Institute for Climate Impact Research.

Gronwald, M., Ketterer, J., Trück, S., 2011. The relationship between carbon, commodity and financial markets: a copula analysis. Econ. Rec. 87, 105–124.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.

Hammoudeh, S., Nguyen, D.K., Sousa, R.M., 2014. What explain the short-term dynamics of the prices of CO2 emissions? Energy Econ. 46, 122–135.

Heston, S., Sinha, N., 2016. News versus sentiment: predicting stock returns from news stories. Financ. Econ. Discus. Ser. 2016, 1–35.

Hintermann, B., Peterson, S., Rickels, W., 2016. Price and market behavior in phase II of the EU ETS: a review of the literature. Rev. Environ. Econ. Policy 10 (1), 108–128.

Hoffman, M.D., Blei, D.M., Bach, F., 2010. Online learning for Latent Dirichlet Allocation. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1. Curran Associates Inc., Vancouver, British Columbia, Canada, pp. 856–864.

Jiao, L., Liao, Y., Zhou, Q., 2018. Predicting carbon market risk using information from macroeconomic fundamentals. Energy Econ. 73, 212–227.

Kearney, C., Liu, S., 2014. Textual sentiment in finance: a survey of methods and models. Int. Rev. Financ. Anal. 33, 171–185.

Keynes, J.M., 2018. The General Theory of Employment, Interest, and Money. Springer.

Koch, N., Fuss, S., Grosjean, G., Edenhofer, O., 2014. Causes of the EU ETS price drop: recession, CDM, renewable policies or a bit of everything?—New evidence. Energy Policy 73, 676–685.

Koenker, R., Bassett Jr., G., 1978. Regression quantiles. Econometrica 33–50.

Krishnamoorthy, S., 2018. Sentiment analysis of financial news articles using performance indicators. Knowl. Inf. Syst. 56 (2), 373–394.

Li, F., 2006. Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports? Available at SSRN 898181.

Li, X., Shang, W., Wang, S., 2019. Text-based crude oil price forecasting: a deep learning approach. Int. J. Forecast. 35 (4), 1548–1560.

Long, J.B.D., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. J. Polit. Econ. 98 (4), 703–738.

Loughran, T., Mcdonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Financ. 66 (1), 35–65.

Loughran, T., Mcdonald, B., 2016. Textual analysis in accounting and finance: a survey. J. Account. Res. 54 (4), 1187–1230.

Loughran, T., McDonald, B., Pragidis, I., 2019. Assimilation of oil news into prices. Int. Rev. Financ. Anal. 63, 105–118.

Marcu, A., Alberola, E., Caneill, J.-Y., Mazzoni, M., Schleicher, S., Stoefs, W., Vailles, C., 2018. 2018 State of the EU ETS Report.

Marcu, A., Alberola, E., Caneill, J.-Y., Mazzoni, M., Schleicher, S., Vailles, C., Stoefs, W., Vangenechten, D., Cecchetti, F., 2019. 2019 State of the EU ETS Report.

Mendel, B., Shleifer, A., 2012. Chasing noise. J. Financ. Econ. 104 (2), 303–320.

Nguyen, T.H., Shirai, K., Velcin, J., 2015. Sentiment analysis on social media for stock movement prediction. Expert Syst. Appl. 42 (24), 9603–9611.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.

Shapiro, A.H., Sudhof, M., Wilson, D.J., 2020. Measuring news sentiment. J. Econ.

Shiller, R.C., 2000. Irrational exuberance. Phil. Publ. Policy Q. 20 (1), 18–23.

Sinha, N.R., 2016. Underreaction to news in the US stock market. Q. J. Financ. 06 (02), 1650005.

Soo, C., 2018. Quantifying sentiment with news media across local housing markets. Rev. Financ. Stud. 31, 3689–3719.

Sousa, R., Aguiar-Conraria, L., Soares, M.J., 2014. Carbon financial markets: a time–frequency analysis of CO2 prices. Phys. A Stat. Mech. Appl. 414, 118–127.

Stone, P.J., Dunphy, D.C., Smith, M.S., 1966. The General Inquirer: A Computer Approach to Content Analysis. M.I.T. Press, Oxford, England.

Tan, X.-P., Wang, X.-Y., 2017. Dependence changes between the carbon price and its fundamentals: a quantile regression approach. Appl. Energy 190, 306–325.

Tetlock, P.C., 2007. Giving content to investor sentiment: the role of Media in the Stock Market. J. Financ. 62 (3), 1139–1168.

Tetlock, P., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. J. Financ. 63, 1437–1467.

Yang, B., Sun, J., Guo, J., Fu, J., 2019. Can financial media sentiment predict merger and acquisition performance? Econ. Model. 80, 121–129.

You, J., Wu, J., 2012. Spiral of silence: media sentiment and the asset mispricing. Econ. Res. J. 47 (07), 141–152.

Zhu, B., Ye, S., Han, D., Wang, P., He, K., Wei, Y.-M., Xie, R., 2019. A multiscale analysis for carbon price drivers. Energy Econ. 78, 202–216.

Zou, H., Sun, L., 2012. The influence of investor sentiment on stock return and its volatility under different market states. In: 2012 Fifth International Conference on Business Intelligence and Financial Engineering, pp. 337–341.