The Influence of Twitter Sentiment in the EU Emissions Trading Scheme

Peter Deeney,

Dublin City University Business School,

Glasnevin, Dublin 9, Ireland

Mark Cummins,

Dublin City University Business School,

Glasnevin, Dublin 9, Ireland

Michael Dowling,

ESC Rennes School of Business,

2 rue Robert d'Arbrissel, 35065, Rennes, France

Alan F. Smeaton,

Insight Centre for Data Analytics,

Dublin City University, Glasnevin, Dublin 9, Ireland

Abstract

We test sentiment measured from Tweets concerning climate change and the EU emissions market, on high frequency price data. Our first main finding is that changes in the sentiment of Tweets specifically concerned with the EU emissions market predict changes in EUA prices two hours ahead with evidence of bi-directional Granger causality between changes in sentiment and changes in EUA prices. Further, we establish that periods of above (below) average sentiment correspond with periods of high (low) EUA return volatility. These findings show that sentiment does indeed have an influence in the EU emissions market. Our second finding is that while energy commodity prices, particularly NBP gas and to a lesser extent Brent oil, can account for some of the movement of contemporaneous EUA prices they are not useful at predicting these changes. This indicates that the emissions market assimilates new information from the energy market quickly. Our third main finding is that there is no evidence that Twitter sentiment concerning the general topics of climate change and global warming (rather than specifically the EU emissions market) is associated with EUA returns. This indicates that the principal means by which the EU is addressing climate change, namely the use of emissions trading, does not seem to register in the general Tweeting of Europeans about climate change and global warming.

Keywords: behavioural finance, sentiment, EU ETS, market efficiency

JEL Codes:

G13, G14, Q57

1. Introduction

The EU emissions trading scheme (EU ETS) is the EU's flagship mechanism to reduce greenhouse gas output and hence to reduce the effects of climate change. European Union emissions allowances (EUAs) must be surrendered by regulated installations for each tonne of CO_2 or equivalent, emitted. These allowances are traded on the ECX. This market is quite large; during the final year of the 2015 December futures contract there was 25bn traded on that one contract. Research on this topic is informing the formation of other trading schemes around the world most notably the Chinese national emissions trading scheme.

A first contribution of this study is our investigation into the impact of sentiment on the European emissions market, while controling for energy market (i.e. Brent oil, NBP gas, ARA coal) and broad equity market (i.e. FTSE) influences at intra-day frequency. While there have been many papers examining the microstructure of the European emissions markets, including Daskalakis and Markellos (2009), Bredin, Hyde and Muckley (2014), Mizrach and Otsubo (2014), Chevallier and Sevi (2014) and Ibikunle et al. (2015) , this is the first attempt to use intra-day frequency data to examine sentiment effects. We propose a simple model that sentiment is directly related to the price and volatility of European Union Allowance (EUA) futures. This is based on a similar model used in Deeney et al. (2015) where the market sentiment index of Baker and Wurgler (2006) was found to be directly related to the price of crude oil futures . To test the relationship with price, we use both multivariate regression and vector autoregression analyses, while to test the relationship with volatility, we use both GARCH and Threshold GARCH (TGARCH) frameworks under scenarios of high and low sentiment.

The use of social media sentiment to analyse political and financial issues is relatively new in the literature and to date no one has analysed the link between social media and emissions markets. In an unpublished study Rao and Srivastava (2012*a*) and Rao and Srivastava (2012*b*) examine several commodities including oil using Twitter sentiment and Google search volume. They find that there is a high positive correlation between the count of positive and the count of negative Tweets and oil price, which is very similar to the findings presented in Section 5. Sprenger, Tumasjan, Sandner and Welpe (2014) use microblogs to measure market sentiment in the DJIA, while Bollen, Mao and Zeng (2011) [1] use Twitter to predict the Dow Jones Industrial Average and Da, Engelberg and Gao (2015) uses the frequency of Google searches to predict volatility in the the S&P500 and several other US indices. Yang, Mo and Liu (2015) show that the people who form a community by communicating with each other using Twitter send Tweets whose sentiment is predictive of stock markets. Sprenger, Sandner, Tumasjan and Welpe (2014) uses the daily count of Tweets as confirmation that a news event has happened and as a method of finding out precisely when an event happened. Lynn et al. (2015) provide a basis for social media sentiment analysis, suggesting that there are several aspects to interactions on social media. Of interest here are the aspects of identity, relationship, reputation, conversation and sharing. We see that Tweets convey a writer's wish to reveal themselves to some extent, to form relationships with others, to build up their reputation and to share opinions and information. A further examination of the rationale behind sharing commercially valuable information is given by Chen et al. (2014) who examine the Seeking Alpha blog. They list four reasons why writers are prepared to place valuable information on a public forum: the writer gets attention, fame and a following; in the case of the blog Seeking Alpha, posters of messages get paid if people read their blogs; writers get a chance to put their opinions into circulation and perhaps fix errors; and finally, writers get a chance to back up their own positions

[1] Should we include this? @Alan their commercial application didn't work and I think there are questions about their methods now.

so that the market will move in their favour. Sprenger, Sandner, Tumasjan and Welpe (2014) points to Information and Salience Theories to explain why Tweets influence stock prices. The Information theory says that Tweets contain information and hence investors can get this information cheaply to make better decisions. Salience Theory says that Tweets direct investors attention to particular stocks causing buying or selling of these. Both agree that there is a direct effect between Tweets and the financial markets. We therefore have an explanation why people will put commercially valuable information into a public domain, Twitter, and how this might influence EUA prices.

Motivated by this emerging evidence, we investigate if there is a role for social media sentiment in explaining emissions market dynamics. Social media analysis is particularly suited to the European emissions market because of the interplay of financial and political influences in the market as attested to by Benz and Trück (2009), Koch et al. (2014), Zhu et al. (2015) and Deeney et al. (2016). Using sentiment measured from specific Tweets about the EU emissions market we are able to improve on the performance of volatility models estimated on EUA futures prices and also improve short term predictions of EUA futures prices. While the improvements are small, they are established to be statistically significant. We find that there is bi-directional Granger causality between sentiment and emissions allowance prices, and, in a secondary finding,between sentiment and the FTSE.

A second contribution of this study is that we find that information from the energy markets, in particular NBP gas (and to a lesser extent Brent Oil) explains EUA prices but do not predict them. This indiates that the emissions market assimilates information from the energ market quickly. To the best of the knowledge of the authors this is the first time EUA, energy market and stock market data have all been used in such a study at intra-day frequency.

An interesting third discovery is that the sentiment measured from Tweets which

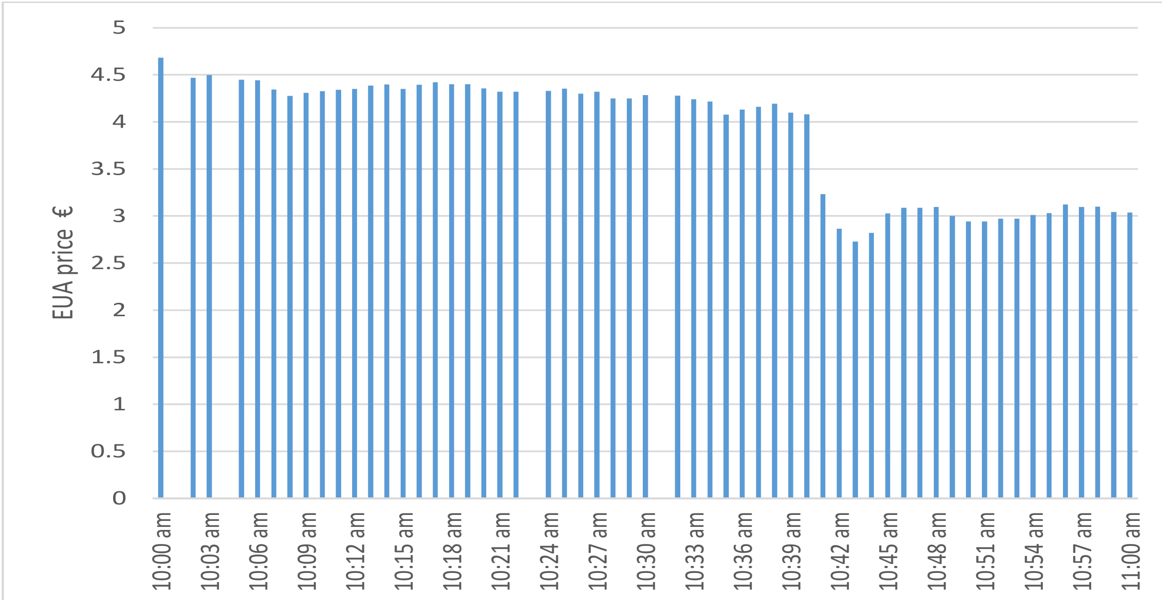


Figure 1: Backloading event prices of EUAs

concern climate change and global warming is not correlated with emission allowance prices. Only sentiment extracted from Tweets specifically concerning the EU Emissions Trading Scheme (EU ETS) is found to a useful factor in describing EUA price dynamics. This suggests that Europeans who are actively tweeting about climate change and global warming are not actively engaged in commentary on esmissions trading.

For our empricial anlaysis, we choose to test the period 17th December 2012 to 16th December 2013. This is the final year of the December 2013 futures contract. During this year the largest ever one-day change in EUA futures prices occured when the price of the EUA prompt December futures contract fell 35% on the 16th April 2013. The sudden collapse in price, which occurred at 10:41am GMT and can be seen in Figure 1. This was due to the European Parliament rejecting a plan which had been intended to support the price of EUAs under a proposal of the European Commission to withold 900 million EUAs from the market and release them at a later date in Phase III of the EU ETS. This process is commonly referred to as “backloading” and was proposed as a way to address the historical oversupply of allowances that resulted from the general overallocation of allowances by Member States to their industries during Phase I and II of the scheme. Backloading was aimed at supporting EUA prices while holding on to the support of EU states who wanted to maintain the supply of EUAs in the long term. We find that that there is a huge spike in both the number of Emission Market sentiment and EUA price returns on the backloading day, and in fact during the same hour.

The preliminary stage of the sentiment measure construction involvesan initial scoping test to identify which search terms are useful in identifying Tweets concerning climate change, global warming and the European emissions market. After manually reading large samples of Tweets selected by these search terms and examining them for subject accuracy, 17 search terms were used for the investigation[2]. The search, conducted through DataSift[3]. A more specific list of 5 search terms[4] yielded 20,883 tweets which are used to produce the Emissions Market senitment measures from which the main results of this investigation follow. One of the most interesting findings of the investigation is that the Climate Change sentiment measures are not useful in explaining or predicting emission allowance prices but the Emissions Market sentiment measures are useful, suggesting a lack of interest in the EU ETS. In addition to searching for particular search terms other criteria are used to select the Tweets to ensure subject accuracy. The Tweets are from European time zones and are written in English as detected by DataSift's language detector. The measures of sentiment are developed from the score assigned to each individual Tweet by DataSift. The positive and negative sentiment scores are treated separately and are not added together to produce an over-

[2]The terms used for the Climate Change Sentiment were: backloading, carbon market, carbon price, carbon trading, climate change, CO2, drought, emission, EU ETS, flood, fossil fuel, geothermal, GHG, global warming, greenhouse gas, renewable and UNFCCC

[3]http://www.datasift.com/ This leading supplier of news and media analytics based in California USA, provided Tweets with individual sentiment scores for the period 17th Dec 2012 to 16th Dec 2013.

[4]The terms used for the Emissions Market sentiment were: backloading, carbon market, carbon price, carbon trading and EU ETS

all sentiment score for a group of Tweets during the observation interval because the literature suggests that positive and negative sentiments behave differently, see Soroka (2006), Sprenger, Sandner, Tumasjan and Welpe (2014) and Akhtar et al. (2013). In addition to the sum of the sentiment scores of the Tweets a simple count of the number of positive and the number of negative Tweets during the observation intervals is used as a robustness test which avoids over-reliance on the scaling accuracy of the sentiment scores. Thus there were four initial sentiment measures based on the positive and negative scoring of each Tweet, and the count of positive and negative Tweets. A fifth measure was the count of all Tweets gathered, this allows a test of whether the sentiment information is useful compared with a simple count. Following Mitra, Mitra and Dibartolomeo (2009) sentiment impact is calculated for each sentiment measure as the weighted sum of the senitment measure where the weighting decreases exponentially. This allows the sentiment impact to decrease with time as would be expected.

The EU ETS is the largest example of a cap and trade emissions market in the world, it is the principal means by which the EU addresses climate change and is economically important for electricity and other energy costs in the EU. Research in this field influences the construction of a national trading scheme in the world's most populous nation and second largest economy, China as well as many other countries. Our principal discovery is that we find that sentiment does explain and even predict EUA prices and that indeed there is bi-directional causality between changes in sentiment and changes in EUA price. Furthermore we see that sentiment influences the volatility of EUA returns. These results are found while controlling for the influence of oil, coal, gas and stock index prices on EUAs; the use of high frequency data for this analysis is an addition to the literature. Secondly we find that while the energy market is able to explain some of the changes in EUA returns it cannot predict them; this indicates that the emissions market can assimilate energy market information very quickly and efficiently. Thirdly while find that the Emissions Market sentiment impact is explains some of the EUA price and volatility, the general Climate Change sentiment impact is not connected to EUA behaviour, which suggests that Europeans who Tweet about climate change are not concerned with the EU ETS.

The rest of the paper is arranged as follows in section 2 we describe the Twitter data and in section 3 we describe the EUA, oil, coal, gas and FTSE high frequency price data. In section 4 we describe the statistical models and we present results in section 5. Section 6 concludes.

2. Twitter Data

In this section the methods for selecting Tweets, gathering sentiment scores for individual Tweets, combining these scores into sentiment measures and calculating sentiment impact are described. This entire process is carried out twice, once for general climate change sentiment and once for specific emissions market sentiment. Tweets are selected by searching for particular words or search terms occurring in the text of all the Tweets posted from 17th December 2012 to 16th December 2013. Location and language are used as additional selection criteria. The sentiment analysis ustilised is that provided by DataSift[5], who provide a score for each measurable Tweet. If positive sentiment is detected the score is a number between 1 and 20 indicating the intensity of the positive sentiment, if the sentiment is detected as negative the score is a negative number between -1 and -20. It is important to treat negative and positive sentiment separately as the literature indiates that they do not simply cancel each other, see Soroka (2006), Sprenger, Sandner, Tumasjan and Welpe (2014) and Akhtar et al. (2013). Sentiment time series are constructed to describe the sentiment on a minute by

[5]See http://www.datasift.com/. This leading supplier of news and media analytics based in California, USA supplied Tweets with sentiment scores for the period 17th Dec 2012 to 16th Dec 2013.

minute basis, these are aggregated later into obervation intervals of length m minutes so as to be compared with the time series of EUA prices and control variables. The use of intra-day sentiment data is one unique element of this investigation. Four sentiment measures are constructed over the 525,600 minutes covering the period under investigation. The four one-minute-frequency time series are respectively based on (i) the sum of the positive scores during each minute, (ii) the sum of the negative scores during each minute, (iii) the count of the number of Tweets containing positive sentiment during each minute and (iv) the count of the number of Tweets containing negative sentiment during each minute. The latter two series based on Tweet counts reduce our reliance on the accuracy of the DataSift sentiment algorithm. A fifth measure that we consider is the count of the total number of Tweets during each minute, irrespective of whether these Tweets had measurable sentiment or not. They are considered to allow a test of the efficacy of the sentiment measure compared with a simple count of Twtiter traffic.

In order to create a simple model of the behaviour of sentiment we follow the method of Mitra, Mitra and Dibartolomeo (2009) and Yu, Mitra and Yu (2013) which allows the sentiment associated with a particular Tweet to remain effective for period after the Tweet was posted, and for its impact to decrase with time. The details of these procedures are given below.

2.1. Selection of Tweets

To select the Tweets for the analysis an initial scoping list of 44 words and phrases are used as search terms, see Table 1. These terms concern climate change, global warming and the emissions market, and are collected from the indexes of several published books (Chevallier (2011*a*), Stern (2006), Serletis (2007), Kaplan (1983), Ellerman, Convery and De Perthuis (2010) and Richter (2010)). The terms are: *backloading, biofuels, biogas, biomass, cap and trade, carbon, clean tech, climate, CO2, dioxide, drought, electricity, emission, emitter, energy market, environment, EU ETS, EU Parliament,* *EUETS, flood, fossil fuel, geothermal, glacier, global warming, greenhouse gas, hydrocarbons, hydroelectric, ice cap, IPCC, Kyoto Protocol, methane, pollution, power plant, power sector, renewable, sea ice, sea level, smelting, sustainab, tar sands, trading, UNFCCC, warming, wave energy and wind turbines.* Random sample of 100 Tweets found by each of these words are manually checked for subject accuracy. It is found that many search terms produce Tweets which are not intended. For example *"IPCC"* selects many Tweets concerned with the Intergovenmental Panel on Climate Change but also produces Tweets concerned with the Independent Police Complaints Commission. It is also found that *"carbon"* produces Tweets concerning greenhouse gases as well as carbon steel, carbon fibre and the description of a colour. Those words and phrases which produce Tweets which are at least 70% accurate for subject are used to produce the climate change list of 17 words namely: *backloading, carbon market, carbon price, carbon trading, climate change, CO2, drought, emission, EU ETS, flood, fossil fuel, geothermal, GHG, global warming, greenhouse gas, renewable and UNFCCC.*

In addition to searching for Tweets containing any of these 17 search terms, Climate Change Tweets are selected to come from Europe and to be written in English. The geographical origin of the Tweets was determined by the timezone in the Tweet metadata, and the language was determined by the language detection system of Datasift. The geographical restriction is to ensure subject accuracy. The restriction to English is to ensure that the authors may check the subject accuracy of samples of the Tweets. A further restriction is imposed so that Tweets may only be concerned with the emissions market and not generally concerned with climate change or global warming. This produced the Emissions Market list in Table 1. In total, 1,522,562 Tweets concerning the topics of climate change, global warming, and emissions markets formed the source for the Climate Change Sentiment Measures. The smaller set of 5 search terms, namely *backloading, carbon market, carbon price, carbon trading and EU ETS* are specifically

	Terms used for Tweet Search
Initial Scoping	backloading, biofuels, biogas, biomass, cap and trade, carbon, clean tech, climate, CO2, dioxide, drought, electricity, emission, emitter, energy market, environment, EU ETS, EU Parliament, EUETS, flood, fossil fuel, geothermal, glacier, global warming, greenhouse gas, hydrocarbons, hydroelectric, ice cap, IPCC, Kyoto Protocol, methane, pollution, power plant, power sector, renewable, sea ice, sea level, smelting, sustainab, tar sands, UNFCCC, warming, wave energy and wind turbines
Climate Change	backloading, carbon market, carbon price, carbon trading, climate change, CO2, drought, emission, EU ETS, flood, fossil fuel, geothermal, GHG, global warming, greenhouse gas, renewable and UNFCCC
Emissions Market	backloading, carbon market, carbon price, carbon trading and EU ETS

The table presents the search terms used for the Initial Scoping, Climate Change and Emissions Market Tweet searches. The Initial Scoping set of 44 terms were used to verify search term accuracy. Adjustments were made so that the set of search terms in the Climate Change Sentiment and Emissions Market lists were at least 70% accurate for subject when random samples of 100 Tweets for each terrm were checked.

Table 1: Summary of search terms for initial search, climate change sentiment and emissions sentiment

related to the emissions market and returned 20,883 Tweets. These were used to produce sentiment measures for what we term the Emissions Market Sentiment. The sentiment measures from these smaller set of Tweets are found to be very useful in predicting and explaining the level and volatility of EUA returns, whilst there is no evidence from correlation or regression tests that the Climate Change sentiment measures are associated with EUA prices. Thus from this stage onwards in the investigation we only concern ourselves with the Emissions Market Sentiment. Discussion of this choice is found in Section 2.4.

2.2. Verifying the Subject Matter of the Emissions Market Tweets

A very useful aspect of using Tweets to measure sentiment is that the Tweets can be read individually to check for accuracy of subject matter, this is the reason only English language Tweets are chosen. In practice it was not possible to check each Tweet but samples of 100 of each search term were tested for accuracy. In order to check the accuracy of the Emissions Market Tweets we follow Sprenger, Sandner, Tumasjan and Welpe (2014) who use the number of Tweets to identify the timing of actual news events. Similary we verify that the largest spikes in the number of Emissions Market Tweets correspond to actual events in the EU ETS. This verification is carried out using Lexis Nexis to search for articles in European newspapers containing the phrases "Emission Allowances" or "EU ETS" or "Carbon Emissions". As these stories are written by professional journalists and selected by professional editors we assume that they are relevant to events in the EU ETS. If the Tweets are concerned with events in the EU ETS then we would expect that large numbers of Tweets would happen when there are important events in the EU ETS. On days on which there is most media interest in the emissions market, we see that the number of Tweets on these days is high. It might be argued that the similarity between the numbers of news stories and Tweets is due to a day of the week effect or a general trend influencing both news stories and Tweets. There is a very strong day of the week effect in the number of both Tweets and newspaper stories released concerning the emissions market, but there is no evidence of a significant trend over the year. (This is somewhat unusual given the general increase in Twitter activity, however the investigation deals with only a tiny proportion, 0.0008%, of the total number of Tweets[6].) We control for the day of the week effect by calculating the residuals in a multivariate regression using the day of the week and a trend as independent variables to explain the number of Tweets or news

[6]http://www.internetlivestats.com/twitter-statistics/

stories each day. The residuals are therefore the number of excess newspaper stories or Tweets above what would be expected due to any day of the week effect or trend. The 20 days with the largest excess numbers of news stories are presented in Tables 2 and 3. The dates are sequenced in decreasing order of number of excess newspaper stories, it is seen that this corresponds almost exactly with the sequence of the largest numbers of days with very high excess numbers of Tweets. It is also seen by reading the new stories and Tweets that on these days there were highly significant events for the EU ETS. This demonstrates that the Emissions Market Tweets are verifilably concerned with the EU ETS based on the origin of the search terms producing these Tweets and based on the distribution of the numbers of Tweets coinciding almost exactly with important EU ETS events. This study adds to the literature by using the greater granularity of the intraday Twitter information to examine at what specific time within the day particular events happened, not just on which days.

2.3. Sentiment Scores and the Calculation of Sentiment Impact Time Series

Having carried out the previous tests which indicate that the Emissions Market Tweets are correctly associated with the emissions market, the next task is to produce high frequency time series so that the sentiment of these Tweets may be compared with the high frequency market time series, i.e. EUA, oil, coal, gas and the FTSE. In our sample of 20,883 emissions market Tweets, 3,777 are deemed positive, 3,756 are deemed negative and 13,351 unclassified Tweets, which include those that contain factual statements rather than opinionated statements, as well as those whose sentiment could not be evaluated. The accuracy of the DataSift sentiment algoroithm has been attested to by Parameswaran et al. (2013). Later in this study, we find that the count of Tweets is not as useful as the sentiment measures, indicating that the sentiment analysis adds value in terms of information.

DataSift assigns a integers between -20 and +20 to each Tweet using their proprietry

Date	Excess News Stories	Excess Tweets	Event / Headline
4th July 2013	62.9	147.4	Poland's veto on backloading rejected in EP
16th April 2013	62.7	1,475.7	Backloading Rejected by European Parliament (EP)
17th April	54.1	367.4	Backloading Rejected by European Parliament (EP)
3rd July 2013	49.1	992.1	Backloading Accepted by EP to be put to individual states
19th February 2013	39.7	337.7	The International Emissions Trading Association supported the European Commission's backloading proposal
1st July 2013	32.7	122.0	Large CDM project, France require more electrical power, UK generation cost of onshore is below all but gas
11 December 2013	31.1	-32.9	Campaignere call to maintain carbon targets, EP tries to rescue carbon credit system, Poland looses backloading battle, tough emissions targets must stay
8th November 2013	29.5	121.4	EU freezes vast raft of carbon credits in bid to relaunch EU ETS; EU approves measures aimed at sustaining local carbon market; breakthrough in EU carbon talks
1st April 2013	21.7	-26.0	Altmaier (German government minister) flags new CO_2 limits; CO_2 tax up č18 per tonne in 2015; New energy tax will force thousands into poverty;
19th June 2013	20.1	391.1	EP votes to freeze the number of permits auctioned; Shenzhen starts the first of seven Chinese ETS and Kazakhstan plans a national ETS

In order to identify important events in the emissions market we count the number of news stories and Tweets per day. The table shows the numbers of news stories and Tweets more than exptected compared with a day of the week and trend model.

Table 2: Excess News Stories and Tweets (1-10) per day for the EU ETS

Date	Excess News Stories	Excess Tweets	Event / Headline
15th April 2013	19.7	699.0	EU ETS faces crunch vote
21st October 2013	18.7	-10.0	Solving climate change; new nuclear power plant in France; Germany lacks political will for CO_2 tax.
22nd May 2013	18.1	-15.9	Aviation ETS Chinese and Indian companies face fines; EU ETS schme applications open; EU summit set to turn climate agenda upside down; nine member states (of EU) support quota freeze;
1st October 2013	17.7	-74.3	Deal on aviation emissions hangs in the balance; focus on China from centralized and coal to distributed and renewable; Green levies on energy bills; ten year wait for carbon price rebound
28th February 2013	15.9	-36.6	ETS tension mounts before vote on quote freeze; ETS environment committee supports but limits allowance backloading
7th May 2013	13.7	155.7	Berlin backs allowance backloading; Merkel speaks with two tongues on climate; tougher taxes urged on emissions;
26th April 2013	13.5	66.4	Barroso urged to take a stand on EU carbon market fix; carbon price taxes UK competitiveness; EP rejects backloading of allowances; EU pledge to save carbon trading deals
9th November 2013	13.4	0.7	Time running out to reach climate deal
20th June 2013	12.9	33.4	EP to support backloading? ; EPP (European People's Party) can accept strictly circumstribed allowance backloading; EP votes to prop up EU's ailing carbon market
2nd July 2013	384.7	12.7	Zombie carbon markets to be shocked back to life; ETS tension mounting ahead of new EP vote on allowance backloading; mixed far curves (futures market) as region awaits CO_2 vote.

Table 3: Excess News Stories and Tweets (11-20) per day for the EU ETS

sentiment detection algorithm where a positive number indicates a Tweet with positive sentiment and a negative score indicates a Tweet with a negative sentiment. We initially construct four one-minute-frequency time series of sentiment comprising, respectively, (i) the sum of the positive scores during each minute, (ii) the sum of the negative scores during each minute, (iii) the count of the number of Tweets containing positive sentiment during each minute and (iv) the count of the number of Tweets containing negative sentiment during each minute. We add a fifth series, the sum of the number of Tweets per minute, so as to test for any difference between the measures of sentiment and the measure of Twitter traffic intensity. The sums of positive and negative sentiment scores, the counts of positive and negative Tweets and the total number of Tweets per minute are not immediately useful because there are many zeros in these series due to the fact that there are much fewer Tweets than minutes. There is of course a more fundamental problem, namely the sentiment of the market one minute after a Tweet has been posted cannot reasonably be considered to return to zero. In order to model sentiment more realistically we calculate sentiment impact following Yu, Mitra and Yu (2013) and Mitra, Mitra and Dibartolomeo (2009). We set the parameters so that the impact of the senitment measure decreases during every minute becoming negligible (1% of original impact) after a set number of days termed the 'decay length'. We define

$$SentimentImpact_t^{Pos,Sum} \equiv \sum_{i=0}^{t-D} Sent_{t-i}^{Pos,Sum} e^{-ri} \tag{1}$$

$$SentimentImpact_t^{Neg,Sum} \equiv \sum_{i=0}^{t-D} Sent_{t-i}^{Neg,Sum} e^{-ri} \tag{2}$$

$$SentimentImpact_t^{Pos,Count} \equiv \sum_{i=0}^{t-D} Sent_{t-i}^{Pos,Count} e^{-ri} \tag{3}$$

$$SentimentImpact_t^{Neg,Count} \equiv \sum_{i=0}^{t-D} Sent_{t-i}^{Neg,Count} e^{-ri} \tag{4}$$

$$SentimentImpact_t^{All,Count} \equiv \sum_{i=0}^{t-D} Sent_{t-i}^{All,Count} e^{-ri} \tag{5}$$

where $SentimentImpact_t^{*,*}$ is the impact of the sentiment measure at minute t, ${Sent^{*,*}}_{t-i}$ is the sum of the sentiment scores or counts during minute $t-i$, (these being the sum of positive or sum of negative sentiment scores, or the count of positive or negative Tweets, or the total number of Tweets); r is the rate of decay of the sentiment impact and is chosen so that $e^{-rD} = 0.01$ when D is the number of minutes in the decay length. Thus we may reasonably say the sentiment of a particular Tweet has a sentiment impact for several days (the decay length) after which its influence is effectively zero. Patton and Verardo (2012) have found a decay length for the effect of news in the equity market of 2 to 5 days, Mitra, Mitra and Dibartolomeo (2009) find a similar effect with a decay length of 7 days and Yu, Mitra and Yu (2013) confirm these time periods. To ensure robustness we use decay lengths from two days to one week; these give similar results.

It is to be noted that there is a very high correlation between pairs of positive and negative sentiment impact series as seen in Table 5. While this would seem to be contradictory for sentiment from a single source, it is not at all surprising for Twitter sentiment. Firstly the sentiment of a Tweet concerning a market event will depend heavily on whether the sentiment holder is in a long or short position. Secondly it is difficult a priori to decide whether a particular event in the market is likely to lead to higher or lower prices. Thus it is not surprising that we observe that positive and negative Tweets occur very close to each other.

These five series of sentiment impacts (we include the count of Tweets as well as the four measures of sentiment) may be aggregated at m minute observation intervals by summing the log returns and sentiment impact measures for these minutes following the pattern of

$$SentImpact(m)_{t/m} = \sum_{i=t-m+1}^{t} SentImpact_i,$$

subject to $\frac{t}{m}\epsilon\mathbb{N}$. This allows a range of granularities for the analysis of prices and sentiment which are chosen so as to suit the EUA data availability. The statistical properties of the resulting series at different observation interval lengths is given in tables 4 and 5.The series themselves and their first differences were found to be stationary using the Augmented Dickey Fuller test. The choice of $m = 60$ for the main results is to suit the EUA data and this choice is described in Section 3, in the testing phase several other values near to one hour frequency are used to ensure robustness of results.

In addition to the five sentiment impact measures we define high and low sentiment periods for each sentiment impact as follows

$$SentHigh_j = \begin{cases} 1, & SentImpact(m)_j > MeanSentImpact(m) \\ 0, & otherwise \end{cases}$$

where $MeanSentImpact(m)$ is the mean of the sentiment impact measure at observation interval m taken over the whole period under investigation; we use $m = 60$ for the reported tests in Section 5. This is particularly useful for the GARCH analysis where it was not possible to achieve convergence using sentiment impact for any of the measures. The practice of dividing time into periods of high and low sentiment has been quite useful, for example the sign of the Fama French RMRF[7], excess return on

[7]the value weighted return on all NYSE, AMEX and NASDAQ stocks minus the one month Treasury bill rate

the market indicates a bull or bear market on a daily basis, Kim, Ryu and S.W. (2014) and Baker and Wurgler (2006) divide their analysis into high and low sentiment. As we treat positive and negative sentiment separately we define sentiment as high if it is above the mean for positive series or below the mean for the sum of negative sentiment. This is repeated using the median in place of the mean yielding similar results.

2.4. Initial Observations of Climate Change and Emissions Market Sentiment

It is seen in Figure 4 that there is no unusual behaviour in the Climate Change sentiment counts or scores on 16th April, 1st February or 3rd July on which the EUA returns have their largest daily changes see Figures 2 and 3. The large negative spike in Climate Change sentiment on 5th December 2013 in Fig 4 is due to public reaction to flooding in the south of England after a winter storm, there is no particularly unusual behaviour in the returns of EUAs on that day. These observations are confirmed by tests which show that correlations between the Climate Change sentiment impact series and the EUA price series are not significantly different from zero. Thus we conclude that the sentiment of Tweets concerning general climate change are not related to EUA price returns. These findings are confirmed by regression analysis.

The smaller set of five search terms specifically for the Emissions Market produces sentiment measures which are seen to have a much clearer connection with the EUA market. In Figure 5 we see that there is a link with EUA returns which have a large negative change on 16th April 2013 and a large positive change on 4th July 2014. This is confirmed later in the statistical tests outlined in Section 4. For the remainder of out investigation we will deal only with the Tweets specifically related to the emissions market.

20 Minute N=7,620	Sum Pos	Sum Neg	Count Pos	Count Neg	Total Tweets
Mean	36.51	-29.94	8.84	6.47	39.60
Max	1036.62	0.00	244.41	108.03	842.07
Min	0.00	-476.77	0.00	0.00	1.33
Median	18.17	-18.83	4.51	4.08	23.63
Std Dev	69.77	37.74	16.33	8.17	59.77
Skewness	7.37	-4.56	7.18	4.92	6.46
Kurtosis	81.68	39.25	79.49	45.28	65.75

Hourly, N = 2,540	Sum Pos	Sum Neg	Count Pos	Count Neg	Total Tweets
Mean	37.23	-30.53	9.02	6.60	40.40
Max	1026.79	0.00	242.27	107.57	837.30
Min	0.01	-476.77	0.00	0.00	1.36
Median	18.52	-19.04	4.57	4.11	23.96
Std Dev	71.41	38.59	16.71	8.35	61.21
Skewness	7.33	-4.54	7.14	4.87	6.41
Kurtosis	79.74	38.34	77.56	43.87	64.11

Daily, N = 254	Sum Pos	Sum Neg	Count Pos	Count Neg	Total Tweets
Mean	36.51	-29.94	8.84	6.47	39.60
Max	1036.62	0.00	244.41	108.03	842.07
Min	0.00	-476.77	0.00	0.00	1.33
Median	18.17	-18.83	4.51	4.08	23.63
Std Dev	69.77	37.74	16.33	8.17	59.77
Skewness	7.37	-4.56	7.18	4.92	6.46
Kurtosis	81.68	39.25	79.49	45.28	65.75

The table presents descriptive statistics for each of the five sentiment impact measures based on the positive and negative Tweet sentiment scores, and the counts of positive and negative Tweets and the total number of Tweets. Impacts are weighted means calculated from the sentiment measure provided by DataSift using $SentimentImpact_t^{*,*} = \sum_{i=0}^{t-D} Sent_{t-i}^{*,*} e^{-ri}$ following Eqn 2, where where $Sent_{t-i}^{*,*}$ is one of the sentiment measures summed during minute $t-i$ these being the sum of the positive scores per Tweet, sum of negative scores per Tweet, the count of positive Tweets, the count of negative Tweets or the count of the all Tweets; r is the rate of decay of sentiment impact and is chosen so that $e^{-rD} = 0.01$ when D is the number of minutes in the decay length. Results are presented for data at 20 minute, hourly and daily frequency.

Table 4: Descriptive Statistics

	Sum Pos	Sum Neg	Count Pos	Count Neg	Total Tweets
Sum Pos	1	-0.75	0.99	0.78	0.90
Sum Neg	-0.74	1	-0.75	-0.99	-0.86
Count Pos	0.98	-0.77	1	0.80	0.92
Count Neg	0.76	-0.99	0.78	1	0.88
Total Tweets	0.89	-0.89	0.91	0.90	1

The table shows the correlations between the five sentiment impact measures for the whole year. The top right shows the results for hourly data, the bottom left shows results for daily data. The negative sentiment impact is recorded as a negative number hence the negative correlation between sum of positive and sum of negative sentiment impact actually means that larger values of positive sentiment occur together with larger values of negative sentiment.

Table 5: Correlation Matrix for Sentiment Impact measures at hourly and daily frequency

Figure 2: Price of EUAs during final year of December 2013 futures contract

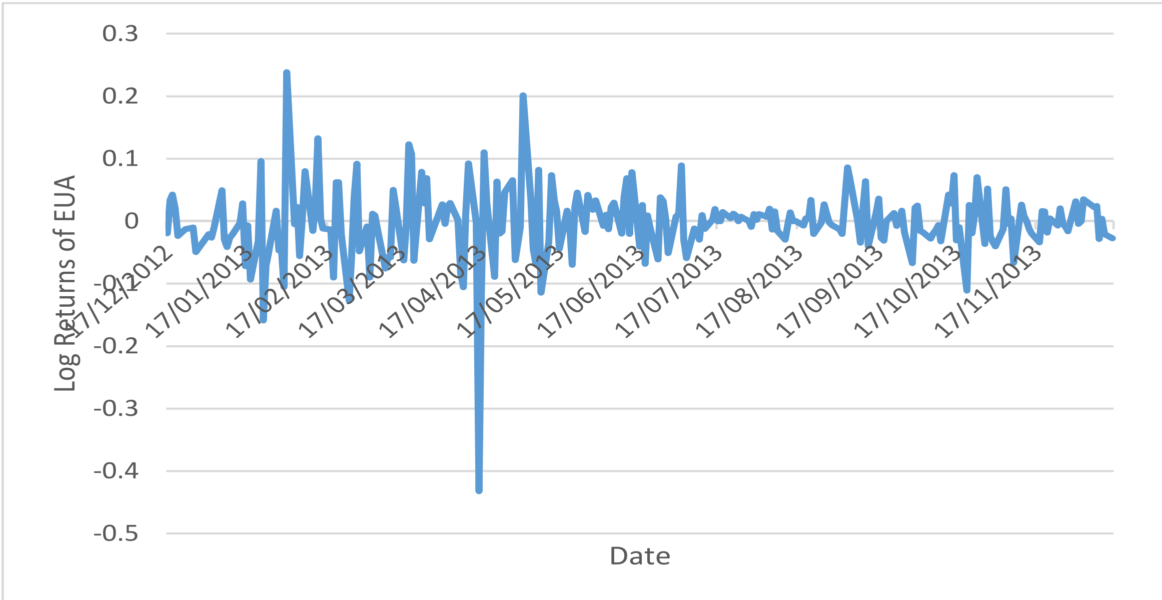


Figure 3: Log returns of EUA from 17th December 2012 to 16th December 2013

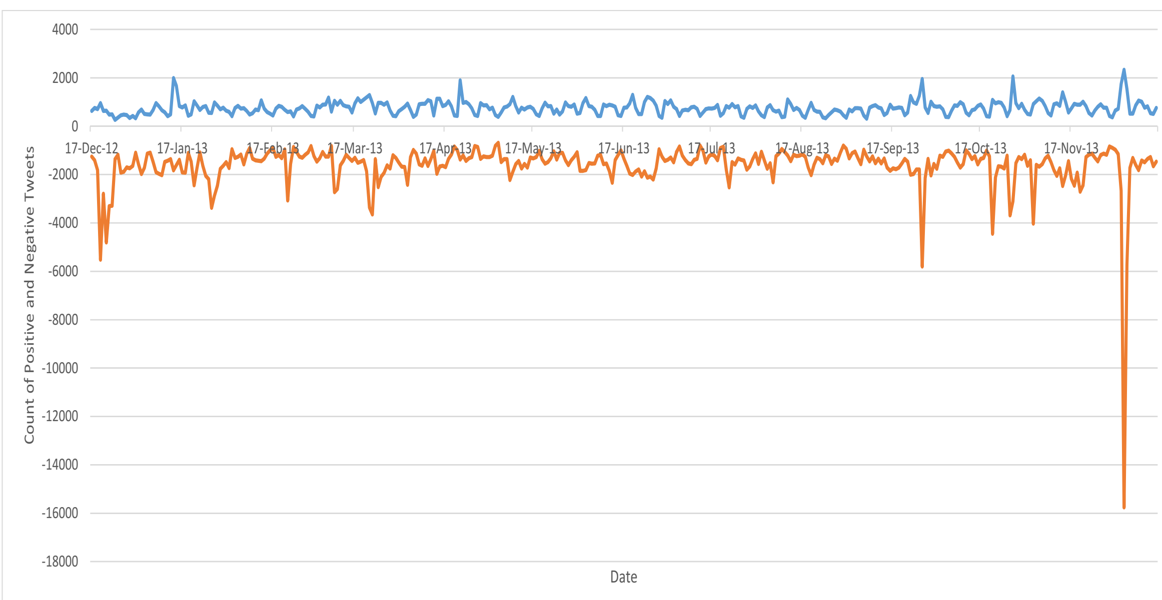


Figure 4: Counts of Positive and Negative Climate Tweets

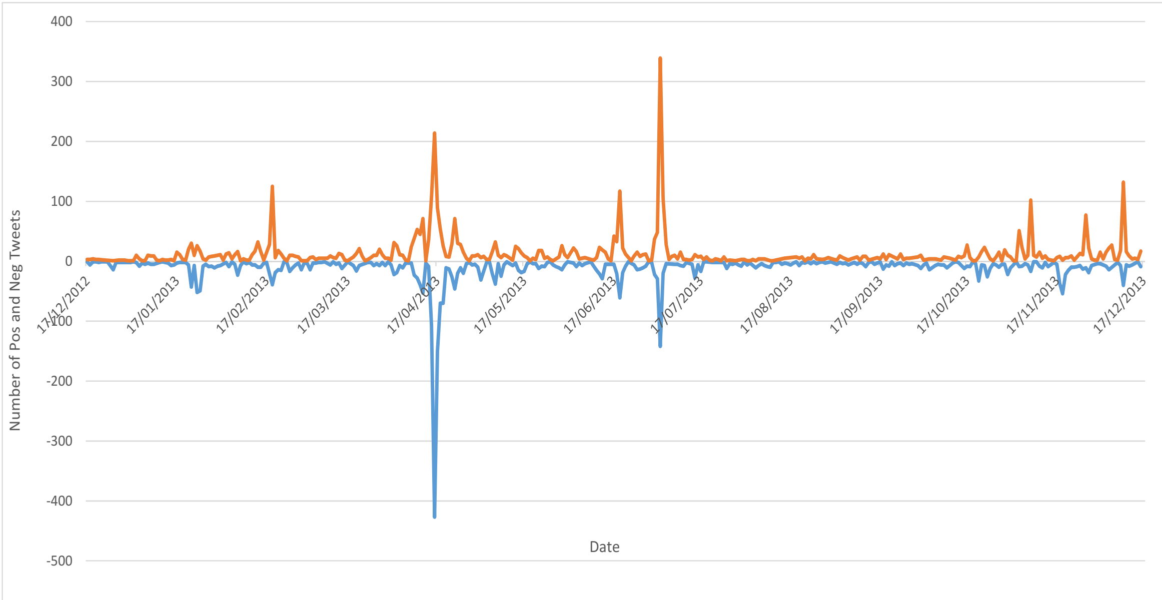


Figure 5: Counts of Positive and Negative Emissions Tweets

3. Emission Allowance, Energy and Market Data

In this section we discuss the choice of the size of the observation interval, the control variables and some possible confounding influences. Following standard practice we use prompt December futures for EU emission allowance (EUA) prices on the ECX as these are the most liquid of the EUA futures contracts, following Mizrach and Otsubo (2014) among others; the data is supplied by the Intercontinental Exchange in London.

3.1. EUA Trading Frequency

Compared with major energy commodities like oil, EUA futures are infrequently traded (see Table 6). It is not the objective of this research to examine the microstructure of the EUA futures market. Very useful work on this has already been carried out by Bredin, Hyde and Muckley (2014) , Chevallier and Sevi (2014), Mizrach and Otsubo (2014) and Ibikunle et al. (2015). In order to determine whether sentiment is a significant driver of returns and volatility it is preferable to avoid microstructure effects. We must therefore choose a suitably large value for m, the observation interval.

Also in order to avoid the effect of bid-ask bounce the EUA price is calculated during every observation interval as being equal to the price during the previous observation interval, if there are no trades, or the mean of the trades during the m minutes of the observation interval weighted by trading volume. [8] This process is also followed for the control variables of prompt month Brent, NBP gas and ARA coal futures as well as the FTSE which are discussed in Section 3.1.1.

When a time series with an observation interval of m minutes is created for EUA December 2013 futures contracts, it is found that there may be many periods with no trading activity, see Table 6. The choice of m, the length of the observation interval measured in minutes, is critical to the number of zeros in the time series. Previous work by Andersen et al. (2001) looking at the Dow Jones found the median duration was 23.1 seconds between trades, and a 5 minute observation interval was used to produce a time series. Similarly, Wang, Wu and Yang (2008) use a 5 minute observation interval for crude oil futures. Both of these markets have far more frequent trades than the EU ETS. In Table 6 we see that if a series of length $m = 5$ minutes is chosen then almost one sixth of these observation intervals would have no trades recorded and hence would have zero as the value for the log return while there would likely be non zero entries for the control variables and for sentiment; this would bias our findings on the possible connections between these variables. The issue is completely avoided by using daily frequency but this would loose much of the information available in the dataset. A reasonable minimal requirement is to require at least 99% of the periods have an EUA transaction. This would be achieved with $m \geqslant 20$.

In addition to avoiding a large number of zeros in the time series we wish to avoid the microstructre effects of the EUA futures market. By by examining serial correlation and order imbalances Chordia, Roll and Subrahmanyam (2005) find that predictive

[8] The calculations are repeated using the median price with no noticeable changes to the results.

Timescale, m	N	Mean No of Transactions	No. of Zeros	% Zeros
1 minute	151,800	2.9	83,765	55.18%
5 minutes	30,360	14.7	4517	14.88%
10 minutes	15,180	29.4	732	4.82%
15 minutes	10,120	44.1	200	1.98%
20 minutes	7,620	58.8	71	0.93%
60 minutes	2,530	176.5	6	0.24%
600 minutes	253	1,764.8	0	0

The table shows the numbers of transactions per observation interval for a series of choices of m the length of the observation intervals in minutes. N is the number of such intervals during the year of our investigation. There were 446,506 transactions for EU emissions allowances on the futures market from 17th Dec 2012 until the expiry of these contracts on 16th Dec 2013. We count only transactions which took place during trading hours of 0700 to 1700 London time and exclude the backloading day 16th April 2013 as it was exceptional.

Table 6: Frequency of Transactions

inefficiencies should not persist beyond 60 minutes on the New York Stock Exchange. This suggests that $m = 60$ would be a safe choice to avoid microstructure effects. A simple but effective way to decide on the length is to select a value of m which reduces serial correlation but retains intra-day frequency. There is very strong negative serial correlation for the first lag of the EUA futures returns when the frequency of the time series is set at $m = 5$, 10, 15 and $m = 20$ minutes; this is expected for high frequency data. There is very little evidence of serial correlation when $m = 60$ minutes from either PACF plots or Durbin Watson tests. We thus conclude that the serial correlation which is a feature of the microstructure of the EUA market, is not present at hourly frequency, however as a robustness check the analysis will be repeated at a range of frequencies near this value.

3.1.1. Energy and Market Controls

Following Bredin and Muckley (2011),Chevallier (2011a), Creti, Jouvet and Mignon (2012), Aatola, Ollikainen and Toppinen (2013), Ahamada and Kirat (2015), Oestreich and Tsiakas (2015) and Koch et al. (2016), and taking into accout data availability we use Brent oil, NBP gas and ARA coal prompt month futures as well as the FTSE for the control variables.

4. Statistical Testing

We wish to test for an association between the carbon market sentiment as measured from Tweets and the returns of EUA futures contracts. As this is the first investigation into the effect of sentiment in the EU ETS we propose a simple model that sentiment drives both price and volatility of EUA prices. A similar direct association was found between sentiment and oil prices in Deeney et al. (2015). In order to investigate the possible links between sentiment and EUA returns we use multivariate regression to test for direct associations, while we use a vector autoregression (VAR) model to examine the effects of lagged variables and the possible Granger causality between sentiment and returns. In order to test the possible links between sentiment and the volatility of EUA returns we use GARCH and Threshold GARCH models. Following standard practice we restrict or attention to trading hours, which are from 0700 to 1700 London time following Zhu et al. (2015).

4.1. Regression Specification

A very simple model is proposed, namely that sentiment is positively related to EUA returns. The effect of the sentiment impact time series are tested using the following predictive regression equation:

$$\Delta EUA_t = \alpha + \beta \Delta SentImpact^{*,*}_{t-pred} + \Delta Controls_{t-pred} + \varepsilon_{t-pred} \tag{6}$$

where ΔEUA_t is the log return of the EUA futures at time t; $pred > 0$, allows for predictive testing and $pred = 0$ allows for contemporaneous testing; $\Delta SentImpact^{*,*}_{t-pred}$ is the first differenced sentiment impact as described in Eqns 2 to 5; $\Delta Controls_{t-pred}$ is

20 Minute N=7,620	EUA	Brent Oil	ARA Coal	NBP Gas	FTSE
Mean x 10 $^{-6}$	14.2	-7.26	-2.72	-1.53	5.71
Max	0.263	0.012	0.176	0.044	0.008
Min	-0.295	-0.001	-0.200	-0.083	-0.008
Median	0	0	0	0	0
Std Dev	0.014	0.002	0.003	0.002	0.001
Skewness	0.465	-0.074	-79.946	-8.388	-0.158
Kurtosis	100.63	6.43	2478.01	505.19	6.53

Hourly, N=2,540	EUA	Brent Oil	ARA Coal	NBP Gas	FTSE
Mean x 10^{-6}	42.68	-21.77	-8.15	-4.58	17.12
Max	0.230	0.015	0.176	0.042	0.013
Min	-0.433	-0.012	-0.200	-0.086	-0.009
Median x 10^{-6}	0	0	0	0	76
Std Dev	0.021	0.003	0.0006	0.003	0.002
Skewness	-2.53	0.122	-4.639	-5.16	-0.11
Kurtosis	97.67	5.29	840.21	178.03	5.496

Daily, N = 254	EUA	Brent Oil	ARA Coal	NBP Gas	FTSE
Mean x 10^{-6}	426.81	-217.66	-81.48	-45.76	171.24
Max	0.268	0.025	0.176	0.046	0.026
Min	-0.448	-0.024	-0.203	-0.078	0.017
Median x 10^{-6}	-461.78	-43.87	0	-152.74	195.99
Std Dev	0.065	0.008	0.018	0.010	0.006
Skewness	-0.564	-0.030	-1.836	-1.214	0.176
Kurtosis	13.46	3.37	91.44	17.12	4.03

The table presents descriptive statistics for Log returns of EUA futures and the control variables of Brent oil, NBP gas, ARA coal and the FTSE. Results are presented for data at 20 minute, hourly and daily frequency.

Table 7: Descriptive Statistics for Log Returns of EUAs and Control Variables

	LnR EUA	LnR Brent	LnR Coal	LnR Gas	LnR FTSE
LnR EUA	1	0.046	0.003	0.069	0.032
LnR Brent	0.116	1	-0.010	0.080	0.181
LnR Coal	0.032	-0.072	1	0.012	-0.022
LnR Gas	0.095	0.098	-0.047	1	0.020
LnR FTSE	-0.043	0.234	0.056	-0.052	1

The table presents the correlations of the log returns of the EUA and control variables. The top right is hourly data, the bottom left is for daily data.

Table 8: Correlation Matrix for EUA and Control Variables

the log return series of Brent oil, NBP gas, ARA coal and the FTSE as described inthe following equation,

$$\Delta Controls_t = \beta_{Brent}\Delta Brent_t + \beta_{NBP}\Delta NBP_t + \beta_{Coal}\Delta Coal_t + \beta_{FTSE}\Delta FTSE_t, \quad (7)$$

where $\Delta Brent_t$ is the log return of the prompt month Brent oil futures,ΔNBP is the log return of the prompt month National Balance Point natural gas price, ΔARA_t is the log return of first month API2 grade coal for delivery to Amsterdam, Rotterdam or Antwerp coal futures and $\Delta FTSE$ is the log return of the FTSE, and the β coefficients are calculated by OLS regression. The data for the control variables are from ICE. The size of the observation interval is measured in minutes and denoted m. This allows the testing to be carried out at a range of frequencies. To avoid the influence of microstructure we choose $m = 60$, a range of values near to this is used as a robustness test. The length of time ahead for predictive tests is given by the product $m.pred$. Results are presented for two hour ahead predictions. A gap of two hours is chosen for two reasons, firstly it ensures that individual trades at time t are at least one hour ahead of the information available at time $t-2$ and so ensures that prediction is a substantial minimal length ahead. This is to avoid the criticism that sentiment is not predicting EUA price movements but merely carrying news information faster than other media and therefore appearing to have predictive value. Secondly we wish to test whether the control variables can predict EUA returns. To do so a short period is chosen so that a negative result may not be due to an over ambitious test. The four sentiment impacts are used, namely the sentiment impact based on the sum of the positive sentiment scores, the sum of the negative scores, the count of the positive Tweets and the count of the negative Tweets. While these last two necessarily loose information by replacing the scaled sentiment measure assigned to each Tweet with a count based measre , it serves as a useful robustness check as it removes reliance on the accuracy of the scaling of the sentiment measure. An additional fifth measure is used which is the number of Tweets per observation interval. This is without reference to their sentiment and allows the efficacy of the sentiment analysis to be tested. In order to compare the relative size of the influences of sentiment and the energy market, variables are standardized. All variables are tested for stationarity using the Augmented Dickey Fuller test.

We run a standard regression as per Eqn 6 for the individual sentiment impacts with the control variables described by Eqn 7. The correlations between positive and negative sentiment measurements are tested at hourly and daily frequency and are presented in Table5. These show that positive and negative sums and counts are strongly correlated, hence these variables are tested separately.

4.2. Additional Concerns

There are three principal concerns, the backloading event, the Samuelson hypothesis and U shaped daily volatility. EUA log returns are seen to be strongly influenced by the Backloading decision of the European Parliament (16th April 2013). The volatility of EUA returns are likely to be influenced by the Samuelson Effect Samuelson (1965) and by the time of dayCont (2011).

On 16th April 2013 there was a narrow rejection of backloading by the European Parliament which caused a huge drop in EUA December 2013 futures prices from 4.76 at the close of busines on 15th to 3.09 at the close of busines on 16th (the drop in the price of the June 2013 expiry futures was even larger but we use December futures as they are the most liquid of the futures). While this single day provides evidence that sentiment from emissions market Tweets and EUA price returns are strongly associated, it may be a single outlier driving the results of regression tests. In order to investigate the un-exceptional behaviour of EUA returns it is prudent to run such tests both with and without the backloading event. This is done by dropping the single day itself from the analysis and by dropping the day and the following four trading days so that any sentiment impact will be removed. There is little difference in the results of these two methods. While temperature has been shown to be an influence on EUA prices Bredin and Muckley (2011), Mansanet Bataler 2007 and Alberola, Chevallier and Cheze (2008) intra-day temperature deviations from seasonal averages were not available.

To avoid the Samuelson effect suggested by Samuelson (1965) ,Andersen et al. (2001), Chang, Daouk and Wang (2009) and Duong and Kalev (2008) we test for effects on volatility up to the end of November 2013 following Chevallier and Sevi (2014) and repeat the analysis while including data up to expiry on 16th December 2013. This did not change the conclusion that there was a highly significant effect of sentiment on volatility.

There is often a high level of volatility due to high frequency of transactions, after opening and before closing of markets each day Cont (2011) and so, to avoid this influencing our tests, we test the effect of the time by allocataing dummy variables for the hour after the first hour on a daily basis. (There is little evidence that time of day has a systematic influence on price level.) Time of the day is shown to influence volatility, but it does not change the conclusion that sentiment has an influence on volatility.

4.3. VAR Model and Granger Causality

Following Sousa and Aguiar-Conraria (2015), Chevallier (2011*b*) and Aatola, Ollikainen and Toppinen (2013) we analyse the three energy prices Brent, NBP gas and ARA coal futures and the EUA futures prices in a dynamic VAR setting to take in to account the possible lagged associations between the EUA prices and the control variables. This will allow any possible serial correlation to be accounted for in the model. In a change of approach we use the FTSE as a measure of economic activity and omit the price of electricity because Aatola, Ollikainen and Toppinen (2013) and Fezzi and Bunn (2009) suggest that electricity price is endogenous. The principal novel aspect of our investigation is the inclusion of Twitter sentiment. In order to make sure that we are not wholly dependent on the accuracy of the sentiment analysis provided by DataSift, in addition to the use of the sum of sentiment scores, both positive and negative, we also use the count of the numbers of positive and negative Tweets, and finally the total number of Tweets. A secondary novel aspect of our analysis is that we use hourly data in place of the usual daily frequency. This high frequency approach gives a new insight into the interactions of the variables in the VAR framework and in the Granger causality testing. The Akaike, Schwarz and Hannan-Quinn information criteria are used to decide the optimal lag length for the VAR and Portmaneau test for auto-correlations.

4.4. GARCH Specification

It has long been the case that sentiment and volatility have been considered to be almost synonymous see Whaley (2000) and Baker and Wurgler (2006). We test this by adding sentiment to a volatility model and measuring any improvement. This method of adding a variable to the variance equation is based on a suggestion by Reider (2009) and similar use by Deeney et al. (2016) and Kumari and Mahakud (2015). GARCH models have been found to be very useful for data which has volatility clustering such as equity markets and commodity futures. We use a standard GARCH(1,1) and a Threshold GARCH(1,1) to test whether the inclusion of sentiment information improves the volatility modelling. We use a binary indicator of high or low level of sentiment which takes the value 1 when the sentiment is higher than the mean and zero otherwise (for the sum of negative sentiment which is measured using negative numbers, we set the dummy variable to 1 when the sentiment impact is below the mean). As a robustness check the analysis is repeated using the median in the place of the mean. We use the usual four measures of sentiment impact based on the count of positive Tweets, the count of negative Tweets, the sum of the positive sentiment scores and the sum of the negative sentiment scores. In addition we add the count of all Carbon Market Tweets.

4.4.1. GARCH(1,1)

We fit the standard GARCH(1,1) model as used by Chevallier (2011*a*), Benz and Trück (2009) and Oberndorfer (2009) and add a sentiment term(s) to test whether this improves the model measuring the improvement with a likelihood ratio test. The equation for the GARCH model is

$$\Delta EUA_t = \mu + \rho \Delta EUA_{t-1} + \epsilon_t,\ \epsilon_t \sim i.i.d.(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma SentHigh_t \tag{8}$$

where ΔEUA_t is the log returns of the EUA price, μ is the drift, $\rho$is the coefficient of first order auto-correlation, $\alpha_0, \alpha_1, \beta, \gamma$ are constants, ϵ_t is the error term with mean zero and conditional variance σ_t^2, and $SentHigh_t$ is one of the binary indicators of sentiment. These take the value 1 when the sentiment impact is larger than average and zero otherwise. (For the sum of negative sentiment impact, which is non-positive, $SentHigh_t$ takes the value 1 when it is below the mean.) We test a series of 4 sentiment impacts based on the count of positive and negative Tweets, the sum of the sentiment of positive and negative Tweets and also use the total number of Tweets. In order to avoid the possible confounding influence of the Samuelson (1965) effect which has been found in EUA return volatility by Chevallier (2009), and the U shaped daily volatility observed in many markets, see Cont (2011), we repeat the GARCH tests without the December data and also control for the hour of the trading day by using dummy variables for the hour after the first hour for each trading day. As a robustness test the analysis is repeated with the control variables of Eqn 7 in the mean equation.

4.4.2. Threshold GARCH

The explanation for different effects of positive and negative sentiment is present in the literature, notably the "negativity effect" mentioned by Soroka (2006), Sprenger, Sandner, Tumasjan and Welpe (2014), Chevalier and Mayzlin (2006) and Akhtar et al. (2013) (see Section 2). Market participants are over-optimistic on average and so respond more strongly to bad news than to the good news they had been expecting Liu et al. (2014) and Feng, Zou and Wei (2011). This behaviour is modelled well by a Threshold GARCH model which allows negative shocks to add to the variance independently from positive shocks. Threshold GARCH is used by Alberola, Chevallier and Chèze (2009), Chevallier (2009) and Byun and Cho (2013) to model EUA price dynamics. Following the same nomenclature as Eqn 8 we test the following Threshold GARCH specification,

$$\Delta EUA_t = \mu + \rho \Delta EUA_t + \epsilon_t,\ \epsilon_t \sim i.i.d.(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 (\epsilon_{t-1}^-)^2 + \beta \sigma_{t-1}^2 + \gamma SentHigh_t \tag{9}$$

where ϵ_{t-1}^- is the value of the previous residuals when it is negative, that is the coefficient α_2 measures the excess volatility due to a negative residual. The same robustness tests are carried out as for the GARCH model regarding the Samuelson hypothesis and U shaped daily volatility.

5. Results

We find there is very strong evidence that in the professionally traded emissions market that sentiment measured from Tweets has an effect on the level and volatility of EUA prices. The sentiment of Tweets concerning the Emissions Market in the EU can improve two hour ahead forecasts of EUA returns and explain volatility, we also find that there is bi-directional Granger causality between sentiment and EUA price. There is very strong evidence from both GARCH and Threshold GARCH models that more extreme sentiment than average does increase the volatility of EUA returns. It is found that the backloading event of 16th April 2013 is so strong an outlier that it can drive regression results on its own, therefore it is omitted from the VAR, GARCH and Threshold GARCH models. These results are now discussed in detail.

5.1. Regression Results

The backloading event of 16th April 2013 has a powerful effect on the analysis and so our decision to test the data with and without that event is shown to be well founded.

We begin by examining the results without the backloading week of 16th April to 22nd April 2013 event as presented in Tables 9, note that similar results follow when only the backloading day 16th April is removed, the five trading days are removed so that any sentiment impact from the backloading event will have dissapated entirely. There is very strong evidence that changes in sentiment as measured by the sentiment impact (see Eqn 2) have a highly significant association with EUA returns, see Table 9. There is less significant evidence that the count of the number of Tweets has this effect as measured by the p-value, the likelihood ratio and the coefficient size. This suggests that the sentiment scores of individual Tweets measured by DataSift does add useful information compared with a simple count of Tweeter traffic. An interesting discovery is that any sentiment increase is associated with price increase (recall that the sum of negative sentiment is a negative number). This is surprising as one might expect rising sentiment to be associated with price rises and falling sentiment to be associated with price reductions. (A similar result from correlation testing is found by Rao and Srivastava (2012*a*) and Rao and Srivastava (2012*b*) for Twitter sentiment and the Brent crude oil.) This may be due to the rather large correlation between positive and negative sentiment generally, as is seen in the correlation results in Table 5. This large correlation between positive and negative sentiment is not entirely unexpected for two reasons. First the opinion of a person posting a Tweet regarding a particular event in the EU ETS may depend on whether they are in a long or short position, so positive and negative sentiment may be triggered in different Tweets by the same event. Secondly individuals may have mixed feelings about a particular event and may differ in their opinions even if they are in a similar long/short position. There is strong evidence that changes in the gas price explain the log returns of EUAs, but there is no evidence of a significant effect from the Brent oil, ARA coal or from the FTSE.

When the backloading event is included in the analysis we find that positive sentiment is highly significantly associated with price increases and negative sentiment is highly associated with price decreases. This may be because the rejection of backloading is easily and immediately interpreted as causing a decrease in EUA price and a generally negative outlook for the of the objectives of the EU ETS.

A predictive model looking ahead by two hours was proposed to test whether sentiment could predict EUA prices. A two hour gap was chosen to ensure that there was at least one hour between the information available and the prediction outcomes, which stretch from 60 minutes ahead to 120 minutes ahead. When we examine two hour ahead predictive regressions without the backloading week, we find that every coefficient of sentiment is significant see Table 10. We observe the same signs for the coefficients as for contemporaneous regressions as seen in Table 9 and see that any increase in sentiment is again associated with an increase in EUA prices. There is no evidence that the control variables can predict EUA prices. When we include the backloading event in the predictive regression we find that the F-tests suggest there is little evidence of any association between the log returns of EUAs and any of the control variables. This indicates that the EU ETS market quickly and efficiently assimilates information from the energy market into EUA prices.

In order to avoid the very high volatility and serial correlation associated with the microstructure of the carbon market we use hourly data. The regression results are robust to selecting the observation interval from the following choices of $m = 40, 50, 60, 75, 100, 120$ minutes while maintaining a 2 hour ahead prediction, these values are chosen as they divide 600 minutes, the length of the trading day. There is an unreported test of the effect of the hour of day which is found insignificant. The usual ADF tests are carried out to confirm the stationarity of the data; Durbin Watson tests and PACF plots show that there is no evidence of serial correlation at $m = 60$ minutes. Breusch Godfrey tests show there is no serial correlation.

5.2. Results of VAR and Granger Causality

Following Sousa and Aguiar-Conraria (2015) we use a VAR model to examine the interactions between the EUA price and the control variables and test for Granger causality. The results of the VAR analysis are presented in Table 11 The data istested for staionarity, the time series were I(0) and the lag length was selected by the Akaike, Schwarz and Hannan-Quinn information criteria. There was some evidence of auto-correlation in the residuals so there is the possibility of missing variables. There is some discussion as to the best model for EUA prices as seen in Koch et al. (2016), so we proceed with caution.

There is remarkable agreement in the Granager causality tests across the four measures of sentiment and the count of Tweets connecting sentiment and log returns of EUAs, see Figures 6, 7, 8, 9 and 10. There is very high significance (p-values below 1%) of bi-directional causality between each of the sentiment measures and EUA prices (except in one case, Sum of Negative, when the p-value is 2.9%). There is also very highly significant evidence of bi-directional Granger causality between these sentiment

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 0	Model 1	Model 2	Model 3	Model 4
	Including Backloading Event N = 2,539						**Excluding Backloading Event N = 2,489**				
Sum Pos		0.048**						0.079***			
Sum Neg			0.184***						-0.077***		
Count Pos				0.034*						0.079***	
Count Neg					-0.188***						0.070***
No. Tweets						-0.079***					
Brent	0.037*	0.036*	0.042**	0.036*	0.042**	0.039*	0.033	0.033	0.032	0.033	0.032
Gas	0.066***	0.065***	0.067***	0.065***	0.067***	0.066***	0.069***	0.069***	0.068***	0.069***	0.068***
Coal	0.024	0.024	0.020	0.024	0.019	0.023	0.017	0.018	0.019	0.018	0.019
FTSE	0.003	0.003	0.002	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003
Log Like	3594.68	3591.75	3550.46	3593.16	3548.72	3586.78	3524.24	3516.46	3516.79	3516.96	3518.03
Ratio T	-	0.0015	0	0.081	0	0.00007	-	0.00008	0.00011	0.00014	0.00042
F test	1.3 x 10^{-3}	2.5 x 10^{-4}	2.7 x 10^{-21}	8.6 x 10^{-4}	2.5 x 10^{-4}	2.9 x 10^{-6}	2.1 x 10^{-3}	5.1x 10^{-6}	6.9 x 10^{-6}	8.1 x 10^{-6}	2.1 x 10^{-5}
R^2	0.7%	0.9%	4.1%	0.8%	4.2%	1.3%	0.7%	1.3%	1.3%	1.3%	1.2%

In this and in Table 10 , the results of OLS regressions are presented. Model 0 does not include any sentiment impact measures, Models 1 to 4 include one sentiment impact measure and Model 5 uses the count of Tweets. The data presented uses contemporaneous log returns of the EUA, Brent, NBP gas, ARA coal and FTSE at one hour frequency from 17th Dec 2012 until 16th Dec 2013. Results are for log returns of all variables except sentiment which is differenced. All variables are then standardized so that meaningful comparisons of the size of the effects of each might be made. The regression equation is

$$\Delta EUA_t = \alpha + \beta_{Sent}\Delta SentImpact_t + \beta_{Brent}\Delta Brent_t + \beta_{NBP}\Delta NBP_t + \beta_{Coal}\Delta Coal_y + \beta_{FTSE}\Delta FTSE_t + \epsilon_t \tag{10}$$

Results are presented with the Backloading event on 16th April 2013 and excluding the trading week from 16th to 22nd April 2013. *, **, *** indicate p-values of below 10%, 5% and 1%.

Table 9: Regression Results for Contemporaneous Data

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	Including Backloading Event N =2,537						**Excluding Backloading Event N = 2,447**					
Sum Pos		0.042**						0.060***				
Sum Neg			-0.054***						-0.083***			
Count Pos				0.044**						0.064***		
Count Neg					0.051**						0.075***	
No. Tweets						0.047**						0.068***
Brent	0.023	0.022	0.021	0.022	0.022	0.022	0.028	0.027	0.026	0.027	0.026	0.027
Gas	-0.004	-0.005	-0.005	-0.005	-0.005	-0.005	0.001	0.001	0	0.001	0	0.001
Coal	0.003	0.003	0.004	0.003	0.004	0.003	-0.0006	-0.005	-0.004	-0.005	-0.004	-0.005
FTSE	0.009	0.009	0.010	0.009	0.010	0.009	0.010	0.010	0.011	0.010	0.010	0.001
Like	3599.95	3597.67	3596.30	3597.47	3596.64	3597.11	3528.77	3524.25	3520.19	3523.67	3521.81	3523.01
R Test	-	0.033	0.007	0.026	0.010	0.017	-	0.0026	0.000034	0.0014	0.0019	0.00069
F test	0.81	0.29	0.13	0.254	0.14	0.20	0.72	0.049	0.0017	0.031	0.00069	0.019
$\mathbf{R}^2$	0.1%	0.9%	0.4%	0.3%	0.3%	0.3%	0.1%	0.4%	0.8%	0.5%	0.6%	0.5%

The Table presents two hour ahead regression results for the following equation.

$$\Delta EUA_t = \alpha + \beta_{Sent}\Delta SentImpact_{t-2} + \beta_{Brent}\Delta Brent_{t-2} + \beta_{NBP}\Delta NBP_{t-2} + \beta_{Coal}\Delta Coal_{t-2} + \beta_{FTSE}\Delta FTSE_{t-2} + \epsilon_{t-2}$$

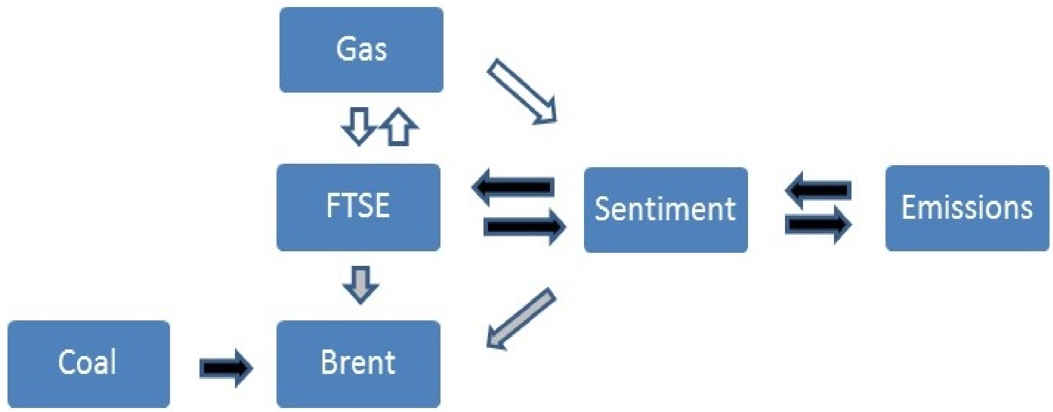
Model 0 does not include any sentiment impact measures, Models 1 to 4 include one sentiment impact measure and Model 5 uses the count of Tweets. The forecast of 2 hours is chosen as it refers to data which is at least one hour ahead and at most three hours ahead of the independent variables (a choice of one hour means the prediction could be very short indeed). Data is similar to that in Table 9 *, **, *** indicate p-values of below 10%, 5% and 1%.

Table 10: Predictive Regression Results

	EUA	Count Neg	Brent	Coal	Gas	FTSE
Constant (x 10^{-3})	0.223	424.743***	0.008	-0.016	0.051	0.011
EUA$_{t-1}$	-0.008	0.879	0.003	0.003	0.005	-0.001
EUA$_{t-2}$	0.038*	-7.023	0.002	0.003	-0.003	0.000
EUA$_{t-3}$	-0.004	-4.370	0.001	-0.002	0.000	-0.002
Count Neg$_{t-1}$	0.00014***	1.077 ***	0.000	0.000	0.000	-0.00002***
Count Neg$_{t-2}$	0.000	-0.066 ***	0.000	0.000	0.000	0.000
Count Neg$_{t-3}$	0.000	-0.045 **	0.000	0.000	0.000	0.000
Brent$_{t-1}$	0.088	2.706	-0.020	0.017	0.020	0.016
Brent$_{t-2}$	0.261 *	-76.068 *	0.014	0.061	-0.012	-0.002
Brent$_{t-3}$	-0.169	-4.274	-0.045 **	0.005	-0.051*	-0.012
Coal$_{t-1}$	0.035	1.550	0.006	-0.011	0.006	0.002
Coal$_{t-2}$	0.034	4.146	-0.008	-0.005	-0.010	0.003
Coal$_{t-3}$	-0.007	0.217	-0.019 **	0.013	-0.003	0.002
Gas$_{t-1}$	0.123	-24.103	0.009	-0.001	-0.014	-0.013
Gas$_{t-2}$	-0.033	37.333	-0.024	0.003	-0.006	-0.032 **
Gas$_{t-3}$	-0.069	15.935	0.020	-0.007	-0.005	-0.008
FTSE$_{t-1}$	-0.170	-116.702 **	0.051**	0.043	-0.017	-0.016
FTSE$_{t-2}$	-0.104	112.423**	0.045**	0.047	-0.029	-0.014
FTSE$_{t-3}$	-0.107	-35.883	0.018	0.130**	-0.010	0.018

The table presents results of the VAR analysis using the Akaike, Schwartz and Bayes information criteria to choose lag length.

Table 11: VAR Results



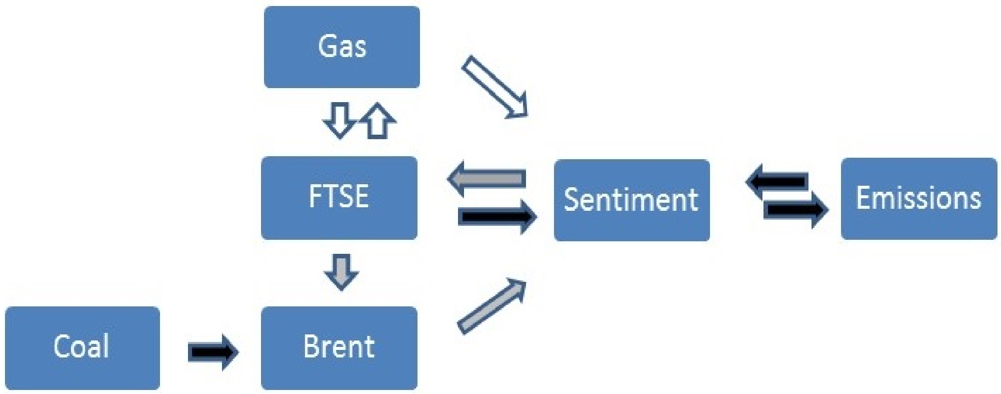
The figure illustrates the Granger causality directions. Black arrows indicate p-values below 1%, grey indicates p-values between 1% and 5%, and white arrows indicate p values between 5% and 10%.

Figure 6: Granger causality directions for Sentiment Sum Positive

measures and the FTSE. There is weaker evidence of bi-directional causality between gas and FTSE, strong evidence of causality from coal to oil and some evidence of sentiment causing oil. Sousa and Aguiar-Conraria (2015) use daily data to find that only a European stock index Granger-causes EUA prices rather than the coal, electricity or gas price. The findings here suggest that this influence from the stock market to EUA price is transmitted by sentiment. Having established a link between sentiment and EUA prices we now present results of the tests of the hypothesis that there is a link between sentiment and the volatility of EUA log returns.

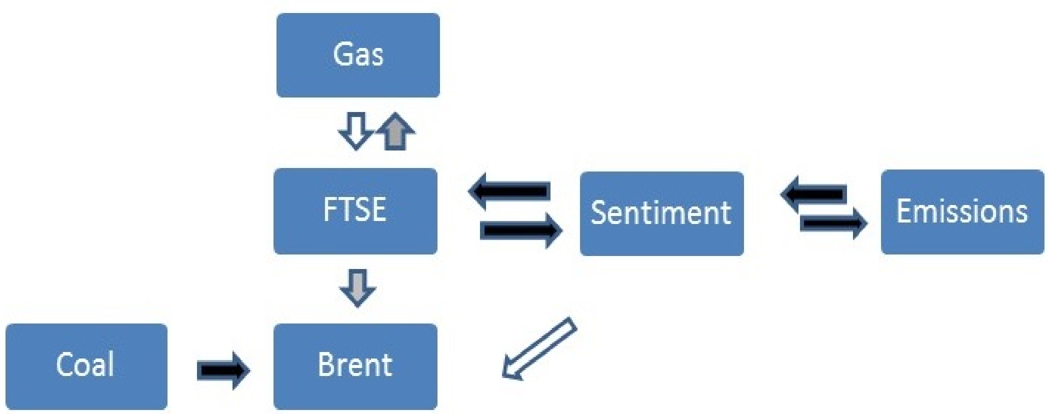
5.3. Results of the GARCH and Threshold GARCH models

We find that all of the measures of sentiment are significant improvements to the standard GARCH(1,1) and Threshold GARCH(1,1) models' ability to explain the variance of EUA returns without sentiment. We measure sentiment here as being either larger or smaller than the mean value of sentiment for the whole year. (In the case of the sum of negative sentiment, which is a negative number, the high sentiment was



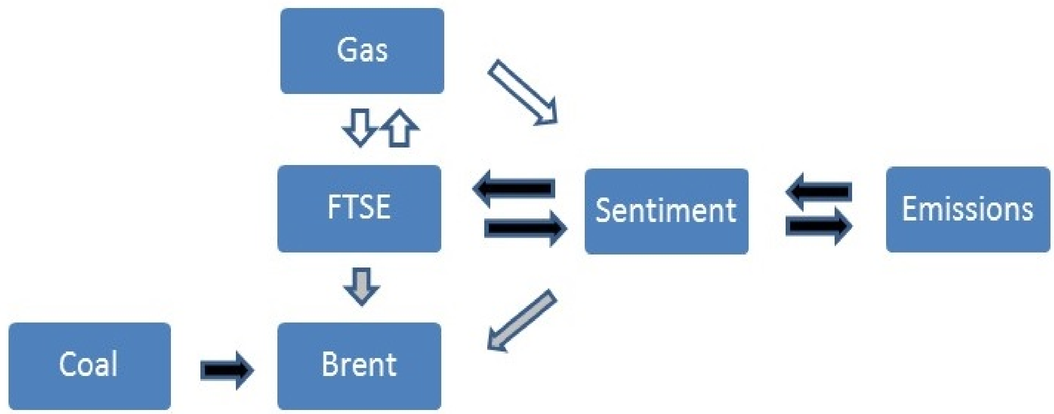
The figure illustrates the Granger causality directions. Black arrows indicate p-values below 1%, grey indicates p-values between 1% and 5%, and white arrows indicate p values between 5% and 10%.

Figure 7: Granger causality directions for Sentiment Sum Negative



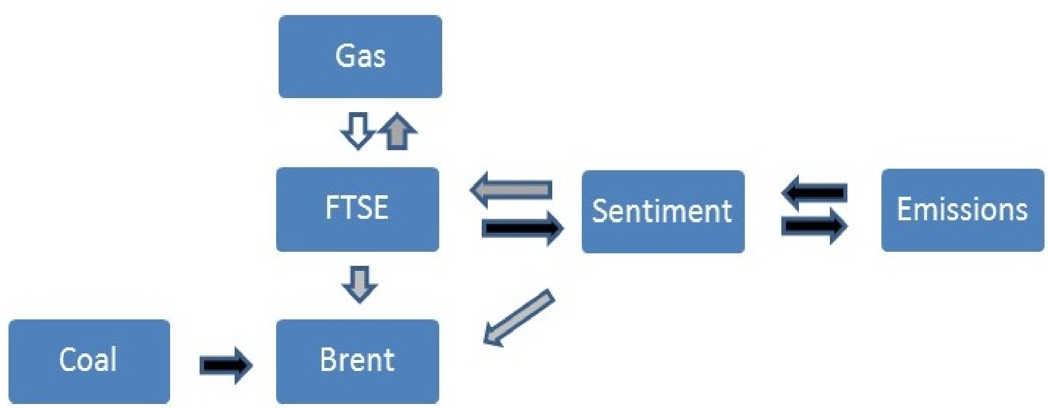
The figure illustrates the Granger causality directions. Black arrows indicate p-values below 1%, grey indicates p-values between 1% and 5%, and white arrows indicate p values between 5% and 10%.

Figure 8: Granger Causality directions for Sentiment Count Positive



The figure illustrates the Granger causality directions. Black arrows indicate p-values below 1%, grey indicates p-values between 1% and 5%, and white arrows indicate p values between 5% and 10%.

Figure 9: Granger Causality directions for Sentiment Count Negative



The figure illustrates the Granger causality directions. Black arrows indicate p-values below 1%, grey indicates p-values between 1% and 5%, and white arrows indicate p values between 5% and 10%.

Figure 10: Granger causality directions for Sentiment Count All Tweets

recorded when the sentiment was below the mean.) The tests were repeated using the median in place of the mean; the same conclusions were reached.

In addition to the usual four measures of sentiment (sum of positive sentiment scores, sum of negative sentiment scores, the count of the number of positive Tweets and the count of the number of negative Tweets) we add the count of any Tweets including those whose sentiment was not measurable. We find that when the number of Tweets is above its mean there is less significant evidence of change to the variance and the size of the coefficient is smaller than for the other measures of sentiment. This would indicate that there is real information in the sentiment scoring method supplied by DataSift. This is the same patterns as is seen in the regression results.

Samuelson (1965) and Carchano and Pardo (2009) suggest that there is an increase in volatility near the maturity date of futures contracts. Chevallier (2011*a*) finds evidence of the Samuelson hypothesis for EUA futures. In order to make sure that this effect is not driving the results, the tests are repeated taking only the data up to the end of trading in November 2013 as done by Chevallier (2011*a*). We find that there is no change to the conclusions and that the Samuelson hypothesis does not interfere with the finding that sentiment as measured from Tweets does have a significant influence on volatility of EUA returns. U shaped volatility in commodity markets during the course of the trading day has been noted by Wolfe and Rosenman (2014) and Batten and Lucey (2010) in commodity markets. We include hour of the day dummies and find that while there is a significant effect on the volatility from some of these dummies, there is still a significant effect from sentiment as measured by any of the sentiment measures. We conclude that higher than average positive sentiment or lower than average negative sentiment as measured by Twitter text is associated with an increase in volatility of EUA returns. This has implications for risk management as well as option pricing.

GARCH Coefficients	No Sentiment	High Sum Positive	High Sum Negative	High Count Positive	High Count Negative	High Number of Tweets
Mean Eqn						
μ(x 10^{-3})	0.656***	0.661***	0.675***	0.658***	0.671***	0.66***
rho	0.023	0.026	0.030	0.026	0.032	0.026
Variance Eqn						
α_0(x 10^{-6})	7.23***	6.06***	6.25***	6.08***	5.82***	6.82***
α_1	0.354***	0.356***	0.359***	0.356***	0.358***	0.354***
β	0.743***	0.737***	0.738***	0.735***	0.738***	0.741***
γ (x 10^{-6})	-	12.8***	6.38***	14.8***	9.41***	4.50**
Log Likelihood	6920.12	6927.94	6923.43	6929.80	6926.45	6921.06
p-value Ratio Test	-	7.7 x 10^{-5}	0.0102	1.1 x 10^{-5}	0.0004	0.171
Dubin Watson	2.07	2.08	2.08	2.08	2.09	2.08

The following GARCH model was fitted using Marquardt with a mean equation $EUA_t = \mu + \rho r_{t-1} + \epsilon_t,\ \epsilon_t \sim i.i.d.(0, \sigma_t^2)$ where the variance of the ϵ_t term is given by $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma Sent_t$. The high sentiment mesured by the sum of the negative scores had the value 1 when the magnitude of the score was larger than the mean magnitude, that is when the negative sentiment was more negative than the mean. In this case the positive coefficient indicates that a more exteme sentiment score is associated with a larger variance.The data here is uses a 3 day delay without the backloading week.

Table 12: GARCH High Sentiment defined as above Mean contemporaneous results

Threshold GARCH Coefficients	No Sentiment	High Sum Positive Score	High Sum Negative Score	High Count Positive	High Count Negative	High Number of Tweets
Mean Eqn						
μ(x 10^{-3})	0.277	0.258	0.290	0.253	0.282	0.268
rho	0.030	0.030	0.034*	0.030	0.036*	0.032
Variance Eqn						
α_0(x 10^{-6})	6.84***	5.44***	6.01***	7.06***	5.60***	6.28***
α_1	0.215***	0.211***	0.217 ***	0.209 ***	0.217 ***	0.213 ***
α_2	0.239***	0.246 ***	0.242 ***	0.247 ***	0.243 ***	0.244 ***
β	0.755***	0.751 ***	0.749 ***	0.749 ***	0.748 ***	0.752 ***
γ (x 10^{-6})	-	13.1***	6.37***	15.2***	9.41***	5.8***
Log Likelihood	6940.19	6949.75	6943.73	6952.04	6946.99	6941.92
p-value Ratio Test	-	1.2 x 10^{-5}	0.0078	1.1 x 10^{-6}	0.0002	0.063
Dubin Watson	2.08	2.09	2.09	2.09	2.09	2.09

The following GARCH model was fitted using Marquardt steps with the mean equation $EUA_t = \mu + \rho r_{t-1} + \epsilon_t,\ \epsilon_t \sim i.i.d.(0, \sigma_t^2)$ where the variance of the ϵ_t term is given by $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2(\epsilon_{t-1}^-)^2 + \beta\sigma_{t-1}^2 + \gamma Sent_t$. Where ϵ_{t-1}^-refers to only the negative values of the residual at time $t-1$, thus it is a measure of the increase of volatility due to a previous negative residual. The figures presented refer to the data without the week of and after the Backloading event and with a delay period of three days, similar results are obtained with just the backloading day being omitted and with a five day delay period (available from authors).

Table 13: Threshold GARCH High Sentiment defined as above Mean

5.4. Discussion

It is clear that there is strong evidence of a link between sentiment and EUA price. There has been previous research from Simon and Wiggins III (2001), Chevalier and Mayzlin (2006), Smales (2015), Akhtar et al. (2013), and Bathia and Bredin (2013) suggesting that negative sentiment is stronger in its effect than positive sentiment, as well as previous research suggesting that sentiment is a contrarian indicator Simon and Wiggins III (2001), and Baker and Wurgler (2006). Here we find that any increase in sentiment is associated with an increase in EUA price. This may well be due to differing opinions of Tweeters to the same events in the EU ETS. While there is statistical significance to the results in this investigation the size of the R^2 for the regressions shows that the models are not very satisfactory. This may be a consequence of using intra-day data which has high levels of volatility compared with daily or monthly data. Another underlying difficulty is that the emissions market is not nearly as liquid as the oil and gas markets, this is unavoidable.

A limitation of this investigation has been that it has only dealt with Tweets in English, this is a limitation imposed by the ability of the authors to personally verify the sentiment and topic accuracy of the Tweets. It would be most illuminating to widen the search to include other European languages.

6. Conclusions

There are three findings in this investigation. Firstly, we find that sentiment as measured from Twitter does have a statistically significant ability to explain and predict EUA prices. Furthermore we find Granger causality in both directions from each of the sentiment measures to EUA price and from these sentiment measures to the FTSE. We find that the coal price seems to be disconnected from the rest of the energy market, this is different from Sousa and Aguiar-Conraria (2015) but may be the result of the use of high frequency data. We also find that high or low sentiment significantly explains some of development of the volatility of the EUA returns.

Secondly, we find that NBP gas significantly explains contemporaneous EUA prices, but none of the energy prices have the ability to predict EUA prices. This indicates that the emissions market is quick to incorporate information from the energy market.

Finally, we find that Twitter sentiment extracted from Tweets concerned with general Climate Change do not have any correlation with the EUA market, only those Tweets specifically concerned with emissions trading have predictive power. This indicates that there is a lack of engagement with the EU ETS. This may be considered surprising as the EU Emissions Trading Scheme (the world's largest emissions market) is the principal means by which the EU aims to reduce greenhouse gas emissions and reduce climate change. It perhaps calls for a greater degree of communications between traders, regulators and the public.

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

Ahamada, I. and D. Kirat. 2015. "The impact of phase II of the EU ETS on wholesale electricity prices." *Revue d'economie politique* 125(6).

Akhtar, S., R. Faff, B. Oliver and A. Subrahmanyam. 2013. "Stock salience and the asymmetric market effect of consumer sentiment news." *Journal of Banking & Finance* 37(11):4488 – 4500.

Alberola, E., J. Chevallier and B. Cheze. 2008. "Price drivers and structural breaks in European carbon prices 2005-2007." *Energy Policy* 36(2):787 – 797.

Alberola, E., J. Chevallier and B. Chèze. 2009. "Emissions Compliances and Carbon

Prices under the EU ETS: A Country Specific Analysis of Industrial Sectors." *Journal of Policy Modeling* 31(3):446 – 462. Climate Change and Energy Policy.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens. 2001. "The distribution of realized stock return volatility." *Journal of Financial Economics* 61(1):43 – 76.

Baker, M. and J. Wurgler. 2006. "Investor sentiment and the cross section of stock returns." *The Journal of Finance* 61:1645 – 1680.

Bathia, D. and D. Bredin. 2013. "An examination of investor sentiment effect on G7 stock market returns." *The European Journal of Finance* 19(9):909–937.

Batten, J.A. and B.M. Lucey. 2010. "Volatility in the gold futures market." *Applied Economics Letters* 17(2):187 – 190.

Benz, E. and S. Trück. 2009. "Modeling the price dynamics of CO2 emission allowances." *Energy Economics* 31:4–15.

Bollen, J., H. Mao and X. Zeng. 2011. "Twitter mood predicts the stock market." *Journal of Computational Science* 2(1):1–8.

Bredin, D. and C. Muckley. 2011. "An emerging equilibrium in the EU emissions trading scheme." *Energy Economics* 33(2):353 – 362.

Bredin, D., S. Hyde and C. Muckley. 2014. "A microstructure analysis of the carbon finance market." *International Review of Financial Analysis* 34:222 – 234.

Byun, S.J. and H. Cho. 2013. "Forecasting carbon futures volatility using GARCH models with energy volatilities." *Energy Economics* 40:207 – 221.

Carchano, O. and A. Pardo. 2009. "Rolling over stock index futures contracts." *Journal of Futures Markets* 29(7):684 – 694.

Chang, C., H. Daouk and A. Wang. 2009. "Do investors learn about analyst accuracy? A study of the oil futures market." *Journal of Futures Markets* 29(5):414 – 429.

Chen, H., P. De, Y. Hu and B. Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media." *Review of Financial Studies* 27(5):1367 – 1403.

Chevalier, J.A. and D. Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43(3):345–354.

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

Chevallier, J. 2011*a*. *Econometric analysis of carbon markets: The European Union Emissions Trading Scheme and the Clean Development Mechanism*. Dordrecht, London: Springer.

Chevallier, J. 2011*b*. "A model of carbon price interactions with macroeconomic and energy dynamics." *Energy Economics* 33(6):1295–1312.

Chevallier, J. and B. Sevi. 2014. "On the Stochastic Properties of Carbon Futures Prices." *Environmental and Resource Economics* 58(1):127–153.

Chordia, T., R. Roll and A. Subrahmanyam. 2005. "Evidence on the speed of convergence to market efficiency." *Journal of Financial Economics* 76(2):271 – 292.

Cont, R. 2011. "Statistical modeling of high-frequency Financial data." *IEEE Signal Processing Magazine* pp. 16 – 25.

Creti, A., P. Jouvet and V. Mignon. 2012. "Carbon price drivers: Phase I versus Phase II equilibrium?" *Energy Economics* 34(1):327 – 334.

Da, Z., J. Engelberg and P. Gao. 2015. "The sum of all FEARS: Investor sentiment and asset prices." *The Review of Financial Studies* 28(1):1–32.

Daskalakis, G. and R.N. Markellos. 2009. "Are electricity risk premia affected by emission allowance prices? Evidence from the EEX, Nord Pool and Powernext." *Energy Policy* 37(7):2594 – 2604.

Deeney, P., M. Cummins, M. Dowling and A. Bermingham. 2015. "Sentiment in oil markets." *International Review of Financial Analysis* 39:179 – 185.

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

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

Ellerman, A.D., F.J. Convery and C. De Perthuis. 2010. *Pricing Carbon, The European Union Emissions Trading Scheme*. Cambridge Univ Press.

Feng, Z., L. Zou and Y. Wei. 2011. "Carbon price volatility: Evidence from EU ETS." *Applied Energy* 88(3):590 – 598.

Fezzi, C. and D.W. Bunn. 2009. "Structural interactions of European carbon trading and energy prices." *The journal of energy markets* 2(4):53–69.

Ibikunle, G., A. Gregoriou, A.G.F. Hoepner and M. Rhodes. 2015. "Liquidity and market efficiency in the world's largest carbon market." *The British Accounting Review* 2015:1 – 17. In Press.

Kaplan, S. 1983. *Energy economics, quantitative methods for energy and environmental decisions*. New York: Mc Graw-Hill.

Kim, J.S., D. Ryu and Seo. S.W. 2014. "Investor sentiment and return predictability of disagreement." *Journal of Banking & Finance* 42:166 – 178.

Koch, N, G Grosjean, S. Fuss and O. Edenhofer. 2016. "Politics matters: Regulatory events as catalysts for price formation under cap-and-trade." *Journal of Environmental Economics and Management* 78:121 – 139.

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

Kumari, J. and J. Mahakud. 2015. "Does investor sentiment predict the asset volatility? Evidence from emerging stock market India." *Journal of Behavioral and Experimental Finance* 8:25 – 39.

Liu, Q., I. Wong, Y. An and J. Zhang. 2014. "Asymmetric information and volatility forecasting in commodity futures markets." *Pacific-Basin Finance Journal* 26:79 – 97.

Lynn, T., S. Kilroy, L. van der Werff, P. Healy, G. Hunt, S. Venkatagiri and J. Morrison. 2015. "Towards a general research framework for social media research using big data." *Working Paper from the Irish Centre for Cloud Cloud Computing and Commerce (IC4) Presented at the Northeastern University International PhD summer school 2015* .

Mitra, L., G. Mitra and D. Dibartolomeo. 2009. "Equity Portfolio Risk Estimation Using Market Information and Sentiment." *Quantitative Finance* 9(8):887 – 895.

Mizrach, B. and Y. Otsubo. 2014. "The market microstructure of the European climate exchange." *Journal of Banking & Finance* 39:107 – 116.

Oberndorfer, U. 2009. “EU Emission Allowances and the stock market: Evidence from the electricity industry.” *Ecological Economics* 68(4):1116 – 1126.

Oestreich, A.M. and I. Tsiakas. 2015. “Carbon emissions and stock returns: Evidence from the EU Emissions Trading Scheme.” *Journal of Banking & Finance* 58:294 – 308.

Parameswaran, A., M.H. Teh, H. Garcia-Molina and J. Wisdom. 2013. “DataSift: An expressive and accurate crowd-powered search toolkit.” *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing* .

Patton, A.J. and M. Verardo. 2012. “Does beta move with news? Firm-specific information flows and learning about profitability.” *Review of Financial Studies* 25(9):2789 – 2839.

Rao, T. and S. Srivastava. 2012*a*. “Modeling movements in oil, gold, forex and market indices using search volume index and Twitter sentiments.”. Accessed on 12th May 2016.

URL: *http://arxiv.org/pdf/1212.1037.pdf*

Rao, T. and S. Srivastava. 2012*b*. “Using Twitter sentiments and search volumes index to predict oil, gold, forex and market indices.”. Accessed on 12th May 2016.

URL: *https://repository.iiitd.edu.in/jspui/bitstream/123456789/31/1/IIITD-TR-2012-005.pdf*

Reider, R. 2009. “Volatility Forecasting I GARCH Models.” *NYU* .

Richter, B. 2010. *Beyond Smoke and Mirrors*. Cambridge: Cambridge Univ Press,.

Samuelson, P.A. 1965. “Proof that properly anticipated prices fluctuate randomly.” *Industrial Management Review* 6(2):41 – 49.

Serletis, A. 2007. *Quantitave and Empirical Analysis of Energy Markets.* Singapore: World Scientific.

Simon, D.P. and R.A. Wiggins III. 2001. "S&P futures returns and contrary sentiment indicators." *Journal of Futures Markets* 21(5):447–462.

Smales, L.A. 2015. "Asymmetric volatility response to news sentiment in gold futures." *Journal of International Financial Markets, Institutions and Money* 34:161 – 172.

Soroka, S. N. 2006. "Good News and Bad News: Asymmetric Responses to Economic Information." *Journal of Politics* 68(2):372 – 385.

Sousa, R. and L. Aguiar-Conraria. 2015. "Energy and carbon prices: a comparison of interactions in the European Union Emissions Trading Scheme and the Western Climate Initiative market." *Carbon Management* 6(3-4):129–140.

Sprenger, T.O., A. Tumasjan, P.G. Sandner and I.M. Welpe. 2014. "Tweets and Trades: The Information Content of Stock Microblogs." *European Financial Management* 20(5):926 – 957.

Sprenger, T.O., P. G. Sandner, A. Tumasjan and I. M. Welpe. 2014. "News or Noise? Using Twitter to Identify and Understand Company-specific News Flow." *Journal of Business Finance & Accounting* 41(7/8):791 – 830.

Stern, N. 2006. *The Economics of Climate Change, The Stern Review.* Cambridge: Cambridge Univ Press.

Wang, T., J. Wu and J. Yang. 2008. "Realized volatility and correlation in energy futures markets." *Journal of Futures Markets* 28(10):993 – 1011.

Whaley, R. E. 2000. "The investor fear gauge." *Journal of Portfolio Management* 26(3):12 – 17.

Wolfe, M.H. and R. Rosenman. 2014. “Bidirectional causality in oil and gas markets.” *Energy Economics* 42:325 – 331.

Yang, S.Y., S.Y. Mo and A. Liu. 2015. “Twitter financial community sentiment and its predictive relationship to stock market movements.” *Quantitative Finance* 15(10):1637 – 1656.

Yu, X., G. Mitra and K. Yu. 2013. “Impact of news on asset behaviour: return, volatility and liquidity in an intra-day setting.” *SSRN 2296855* .

Zhu, B., J. Chevallier, S. Ma and Y. Wei. 2015. “Examining the structural changes of European carbon futures price 2005 to 2012.” *Applied Economics Letters* 22(5):335–342.