Agenda Setting: Twitter Discourse on Climate Change as Signal and Catalyst for ESG Investment

Zhaohan Dong

London School of Economics and Political Science

Author Note

Submitted 2022 as dissertation, any methods or opinions included do not constitute as investment advice.

Abstract

This paper intends to measure the agenda-setting effect of climate change thematic social media on environmentally friendly investment in the equity market. The paper posed the hypothesis of social media's object salience agenda-setting effect can exert influence on the clean, oil and gas, and utilities company stock price returns, and the potential agenda-setting effect between traditional news and social media as a precursor. It analyzed the fitness of using climate change Tweet volume as an additional factor to Fama-French three factor model in explaining portfolio price returns of clean, oil and gas, utilities and green-minus-brown. It also compared the Tweet volume with the green investment index proposed by Brière and Ramelli (2021) and its comparison with Engle et al.'s (2020) content analysis of Wall Street Journal. Finally, it analyzed the correlation between the residuals of news article counts and Tweet volume from AR(1) autoregressive model fitting. The results cannot support the hypothesis of agenda-setting effect on climate change topic from social media to equity market, but are consistent with adaptive market hypothesis and rational inattention. A weak correlation between the AR(1) residuals of news article counts and Tweet volume was found, in support of the agenda-setting effect between traditional news media and social media on climate change topic. The paper discussed the results, and in meta-analysis of related researches, summarized four conditions under which the social media and traditional media affect the equity market.

Keywords: Social Media, Agenda-setting, Climate Change, Adaptive Market Hypothesis, Rational Inattention

Agenda Setting: Twitter Discourse on Climate Change as Signal and Catalyst for ESG Investment

The concern on global warming caused by greenhouse gas emission was first raised decades ago in 1953 (Ungar, 1992). However, for much of the late 20th century, despite the increased governmental awareness and news coverage, global warming failed to gain attention as a "household term" (Schneider, 1989, p. 192; Ungar, 1992). Multiple intergovernmental conferences have been held, and numerous proctocols and frameworks were signed, but no concrete commitments have been taken by companies and individuals alike. Neverthless, Environmental, Social and Governance (ESG) investment, a term raised by UN with a consortium of investment partners, has recently made popular both within the investment community and in the general population (UN Global Compact Initiative, 2004). The prelude to the wax of ESG investment can be traced back to 2007, when a group of Swedish pension funds prompted their bank, SEB, to create financial investments that make positive impact on the world (World Bank, 2018). Working with the World Bank, SEB "linked the dots" between sustainability and financial risk, making ESG investment appealing to all investors, and the first green bond was issued by the World Bank in 2008 (World Bank, 2018).

Given this trend, this paper examines whether the volume of social network posts related to climate change, specifically on Twitter, exhibits statistically significant relationship with stock prices of environment-related sectors, namely clean, oil and gas and utilities. It also examines the relationship between Tweet volume and the green sentiment index of ETF inflow as proposed by Brierè and Ramelli (2021), as well as between Tweet volume and the climate change news article count on Associated Press and New York Times. If a positive relationship between Twitter volume on climate change and green asset return can be identified, it signifies that the asset price and the associated expectation are partially deterimined by the public discourse surrounding a particular issue.

Essentially, this paper aims to discuss whether media can play a key role in binding the "irreconciliable" in "common wisdom" between environment and business through agenda-setting effect, and if a more pronounced discussion of environment in the public online discourse could incentivize environmentally friendly investment (Walley & Whitehead, 1994). If such condition exists, then an increase in Twitter exposure on climate change can become a strategy for the activists and lobbyists to direct fundings to clean company, technology and practices. Furthermore, if the same agenda-setting process exists between traditional media and Twitter, in which traditional media can influence the discourse on Twitter, then the traditional media can also assume a positive role in the environmental protection. Rakowski et al. (2021) have found that social media and traditional media have additive effect in their contribution to the diffusion of information to retail investors.

This approach differentiates itself from various similar researches and accounts of the climate change and investment theme, which center on the news reports of corporate social responsibility (CSR) and individual company performance[1], the action of which might be motivated by reducing negative consumer image through media (Haddock-Fraser & Tourelle, 2010). The concern on the negative company image provides negative rewards, causes aversion for investors and encourages

[1] See Cahan et al., 2015; Lins et al., 2017; McGuire et al., 1988

greenwashing. But such mechanism does not provide positive rewards to investors, with which the investment in climate-positive capital is positively associated with return. This paper attempts to argues that, instead of encouraging investors to divest from polluting companies, a positive influence on clean companies' returns from online discourse on climate change can encourage the investors to increase their investment in clean companies. Furthermore, rather than on an idosyncratic company level, the approach seeks to address media influences on an aggregated market level, so the effect is less concerned with each company's perculiar strategy and more with the overall market sentiment, a characteristic tradition of mass persuation agenda-setting studies (Wanta & Ghanem, 2007).

The paper also contributes to the "Adaptive Market Hypothesis" explained by Lo (2004), which offers an alternative explanation of the market besides the "Efficient Market Hypothesis" raised by Samuelson (1965) and summarized later by Fama (1970). The efficient market hypothesis argues that most agents in the market are rational, and the price of the market is rationally determined through the profit-seeking activity of arbitrageurs whenever the price deviates from the equilibrium. The adaptive market hypothesis states that, in addition to the concept in efficient market hypothesis, while most people are rational, the market does not "incorporate all information rationally and instantaneously", and is contingent on behaviorial bias of individuals (Lo, 2004). Such behavorial bias engender what Shiller (2015) characterizes as speculative bubbles, formed by "feedback loop" of traditional news media.

While the result of this paper failed to find evidence of thematic social media information on

Hypothesis and Background

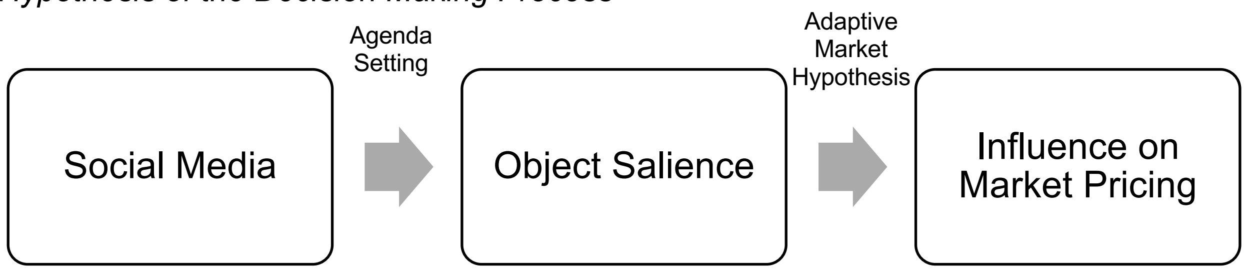
The main hypothesis of the paper is that, the social media exhibits similar agenda-setting attributes as traditional mass media in environmental protection and global warming topics and assert influnce on investments, on both the level of "[transmitting] object salience" and the level of "[transmitting] attribute salience" (McCombs & Ghanem, 2001). In this case, the object salience is more readily observable than attribute salience, since the dataset on social media is large and the measurement accuracy of attribute salience is subject to various factors. The result of the agenda-setting through setting the object of discussion might not only be reflected in public opinion or government policy, as measured by other reseraches, but also the investors' opinion and potentially affect the trading activity and pricing in the equity market.

The social media may also be influenced by the traditional media in providing financial information as noted by Jiao et al. (2016) in their research *Social Media, News Media and the Stock Market*. To capture such effect, this paper also includes news article data from Associated Press and New York Times and analyzes its correlation with social media data.

The sections below will briefly introduce the theoretical frameworks to illustrate the result: the agenda-setting theory, the adaptive and efficient market hypotheses. To elucidate the context surrounding the research, it will also describe the networked social media and financial market, and the current trend in Environmental, Social and Governance (ESG) and green investment. Figure 1 briefly illustrates the pathway from social media to stock market through agenda-setting and adaptive market hypotheses.

Figure 1

Hypothesis of the Decision Making Process



Agenda-setting

The agenda-setting theory was first introduced by McComb and Shaw (1972) in their research on the agenda-setting capacity of mass media in the 1968 U.S. presidential election, although the observation of agenda-setting has been made as early as in 1922 by Walter Lippmann. In the book *Press and Foreign Policy*, Bernard Cohen (1963, p. 13) succinctly described that "[the press] may not be successful much of the time in telling people what to think, but is stunningly successful in telling its readers what to think about." Taking inspiration from Cohen's (1963) work, McComb and Shaw (1972, p. 177) hypothesized that "the mass media set the agenda for each political campaign, influencing the salience of attitudes toward the political issues". They then conducted their research on this hypothesis in the Chapel Hill community around mass media reports of the 1968 presidential election. In their research, a high correlation of +0.967 and +0.979 has been found between voter's independent judgements of what were the important campaign issues and the media's major and minor emphasis on the issues respectively (McCombs & Shaw, 1972, pp. 180–181). From the result, they argued that the thinking of the voters on the relevant political agendas are highly influenced by the media. Although the study did not conclusively prove the agenda-setting theory, as it did not scrutinize individual's throught process in arriving at the conclusion of which political agenda is being emphasized, the exhibited correlation between media political content and voter response is in line with the hypothesis raised by agenda-setting theory (McCombs & Shaw, 1972).

From the discussion of research result, McCombs and Shaw (1972) have concluded that the association between the voters perception of political agenda and the political news coverage can be more readily explained with agenda-setting theory than with selective perception theory. Selective perception theory states that the information decoder, in this case, the readers of news media, will "selectively see and hear based on their needs, motivations, experience, backgrounds, and other personal characteristics" and "project their interests and expectations into communications as they decode them" (Robbins & Judge, 2016). If selective perception effect is indeed stronger than agenda-setting effect, then the voters' perception of agendas will likely to be more influenced by political news akin to their own position, and display stronger correlation with news closer to their own party/candidate preference than with all news. In their Chapel Hill research, "18 out of 24 possible comparisons show voters more in agreement with all the news rather than with news only about their own party/candidate preference" (McCombs & Shaw, 1972). McCombs and Shaw (1972) argued that this particular result rejects the effect of selective perception theory in favor of agenda-setting theory.

McCombs and Shaw's (1972) study further set the framework upon which numerous researches were conducted and the origin point from which the hypotheses on different levels and processes of agenda-setting were constructed and researched (e.g. Beckett, 1994; Cook et al., 1983; Iyengar et al., 1982; Winter & Eyal, 1981). As the number of studies was increased and its subject fields were expanded, two predominant levels of agenda-settings have been proposed: object salience transmission and attribute salience transmission (Ghanem, 1997; McCombs & Ghanem, 2001). The object salience transmission, which is the original agenda-setting theory being researched by McCombs and Shaw (1972) in their 1968 Chapel Hill study, reflects the hypothesis of Cohen (1963) that the media can influence what the audience is thinking about, despite a lack of influence on the audience's approach to the subject. Subject to the scope of our research and available resource, the object salience aspect is more readily researchable as the social media data can be filtered by keywords and counted, and similar researches have been established regarding other topics.

The attribute salience transmission, explicitly linked with the concept of framing, influences how the audience think about the object through the attributes of objects in the framing process, which comprehends cognitive attributes and affective attributes (McCombs, 2005, p. 546; McCombs & Ghanem, 2001, pp. 70–71). This level extends beyond Cohen's (1963) hypothesis, and contributes to how the audience perceives the problem, causes and potential solutions, as identified by Benton and Frazier (1976), as well as the "feeling and tone" toward the object (McCombs, 1992; McCombs & Ghanem, 2001; Patterson, 1993). The media does not only provide an agenda for the information recepient to discuss, but also shapes the through process of the recepient and therefore influences the conclusion of the discussion. A preliminary research using VADER to analyze the sentiments expressed in Tweets containing climate change words was conducted (Hutto & Gilbert, 2014). VADER outputed the sentiment score of each Tweet's text content on a scale between -4 and 4, with negative signs representing negative sentiment, positive signs representing positive sentiment, and zero representing neutral sentiment. The absolute value of the setiment score signifies the strength of the sentiment. The result did not indicate any significant trends (mean = 0.0673, sd = 0.3818). Based on this result, the research did not proceed to measure the attribute salience transmission.

The majority of studies regarding agenda-setting are concerned with the traditional news media toward the public sector and policy making. Nevertheless this research intends to apply agenda-setting concept to social network media toward environmental issues and the private sector in this research. On the agenda-setting effect in addressing environmental issues, Pollach (2014) argued that the news media has significant agenda-setting effect on the environment and pollution topic, because these topics are found unobstrusive by Zucker (1978). People have to rely solely on news to gather information on unobstrusive topics with little to no direct contact with the issues themselves, and such topics are mostly influenced under agenda-setting effect. The effect of media as an intermediate layer on these topics was confirmed by Ader (1995) in her study of agenda-setting for environmental pollution and Olausson (2011) in her study on news media and the perception of climate change among Swedish citizens. The paper also continues on the tradition of mass persuation studies in agenda-setting theory, in which a group rather than an individual from the "societal level" is studied (Wanta & Ghanem, 2007). A notable difference is that, in traditional mass persuation studies, news media sources encode the information and news readers receive and decode the information, whereas in this study, the social media users can encode, receive and decode the information, and both institutional and retail investors may act upon such information in the stock market.

Regarding the approach to agenda-setting investigation, noting factors described by McCombs (1981) and McCombs et al. (1995), Wanta and Ghanem (2007) summarized the previous studies with four factors: single or multiple issue, individual or aggregate data, media content or media exposure and cross-sectional study or longitudinal study. Following such categorization, this research will explore on single issue, aggregate data, media content and conduct a longitudinal study.

Adaptive and Efficient Market Hypotheses

The efficient market hypothesis is the hypothesis that under which, the "security prices at any time 'fully reflect' all information" (Fama, 1970). It was first discussed by Samuelson (1965) that the price of a security follows a random walk and reflects all available information of market particpants on the future expectation of the security's price. The price of the security would fluctuate around the equilibrium and be completely unpredictable when the market is frictionless: all information is available to all market participants and the trading is costless without restrictions (Fama, 1970; Samuelson, 1965). The foundation of the efficient market hypothesis lies upon the principle of supply and demand, which states "the price of any commodity and the quantity traded are determined by the intersection of supply and demand curves" (Lo, 2004). The producer provides the supply curve which constitutes the quantity and price at which she is willing to sell; the consumer provides the demand curve which indicates the quantity and price she is willing to purchase. The equilibrium is reached at the intersection of supply and demand curves, at which the supply and demand are met at the same quantity and price. In market, the equilibrium is reached as the market participants engage in arbitrage, seeking asset mispricings that deviate from the curves. In a frictionless market where investor can act on information instantaneously (efficient market hypothesis condition), no investor can systemically obtain economical gains from actively investing based on available information. To make the condition less restrictive for empirical researches, three levels of tests are formed: strong-form test assumes investors have monopolistic acccess to all information, including private information, and the price reflects all available information; semi-strong-form test assumes the investors have access to all publicly available information, and events will immediately be reflected on pricing; weak-form test assume the price is only determined from historical price information (Fama, 1970). This research assumes the semi-strong-form efficiency of market, because the mispricings still occurs and investors cannot act on private information due to insider trading restrictions.

In his paper to introduce adaptive market hypothesis, Lo (2004) argued if the market is fully efficient in the extreme case, there will be no purpose for active investment, which is commonly practiced in the financial sector, for active investment cannot theoretically achieve economical gains in such condition. Regarding the rational economical agent, Herbert Simon (1955) suggested that the human is bounded by biological limitations and does not possess the physical capacity to carry out completely rational decision computations, and a simplified decision model of the world can introduce discrepancies with the real world. Individuals will make choices that are "merely satisfactory, not necessarily optimal" (Lo, 2004, p. 16). Noting the implication of Simon's work and the gap between it and efficient market hypothesis, Lo (2004, p. 16) suggested that individuals select the best guess of the optimal choice based on past experience, and adapt to changes through positive or negative reinforcement, similar to practices demonstrated in biological evolutionary dynamics. He argued that the sociological and psychological backdrop shall be introduced in addition to the efficient market hypothesis and an inductive approach from observations has to be incorporated with the strictly deductive approach in neoclassical economy (Lo, 2004). Such claim is supported by several emprical researches demonstrating the influence of affect on physiopsychological response and decision of various economic agents (Elster, 1998; Grossberg & Gutowski, 1987; Lo & Repin, 2002; Loewenstein, 2000; Peters & Slovic, 2000).

An important implication from Lo's (2004) adaptive market hypothesis is the more complex market dynamics that allowed existence of "cycles as well as trends, and panics, manias, bubbles, crashes, and other phenomena that are routinely witnessed in natural market ecologies." The market does not follow the "inexorable trend towards higher efficiency" as in efficient market hypothesis, but will see the "creation", "exploition" and "die out" of arbitrage opportunities, resembling the evolution in biological and sociological contexts (Lo, 2004). Whilst the investment decision makers seek to make perfectly rational decisions, and indeed act reasonably rationally for most of the time, they cannot circumvent the psychological tendencies of human. This is where the agenda-setting effect of media can potentially prompt market price movement.

The information in the discovery of arbitrage opportunities can be based on either concrete financial information, or alternative information. Essentially, the investor might perceive the salience of a topic through the agenda-setting effect of social media, and the psychological result of which will influence the decision-making process, and the result could become observable as the price fluctuates.

The Internet, Social Media and Financial Market

The backbone infrastructure that enabled the propagation modes of social media and modern financial market is the Internet. The first precursor of the Internet, ARPANET, was developed in the 1960s and 1970s by the US Defense Advanced Research Projects Agency (DARPA) to connect computers between universities for research projects (Cerf, 1993). The idea is based upon Marill and Roberts' (1966) paper on time-shared computers, that a user of a certain computer installation can access other installations remotely and expect the same functionality across different hardware configurations. In the following decades, other innovations were introduced, such as TCP/IP, DNS and the use of Graphical User Interface (GUI) in accessing websites, all of which has shaped the transition from ARPANET to the World Wide Web, developed by CERN, and the Internet as we come to know today (Elton & Carey, 2013; Robinson & Coar, 2004; Sandvig, 2013). These developments, summarized by Castells (2013, p. 56) as the "technological transformation", enables the transformation in "the organizational and institutional structure of communication", with which "globalization and concentration of media business through conglomeration and networking" and the proliferation of social media occurs. As a result, the interconnectivity and number of users have seen a continuous increase since the 1990s (Wellman et al., 1996).

Built upon the super-infrastructure of the Internet, social media websites and Electronic Communication Networks (ECNs) are the dominant infrastructures in the areas of one-to-many communication and financial transactions, which inherit the "links with conventions of practice" and "embodiment of standards" attributes common to all infrastructures as described by Star (1999). These two infrastructures defined the arena in which the subjects of our research found themselves in, with absence in hierarchy, extension in activity hour and increase in information exchange efficiency: activists and citizens can engage in extended “permanent campaigns” and discussions with no fixed hours and locations; traders engage in off-exchange transactions without intermediaries, which introduced after-hour trading and attracted more retail traders (Bennett, 2003; Castells, 2011a, 2011b; Fan et al., 2000; Wellman et al., 1996). Internet has moved the financial market closer to a frictionless market, and hastened the information propagation process.

In this networked environment, several researches have hypothesized that the trading behaviors on the financial market, especially those exhibited by individual retail investors, are influenced by sentiment and attention rather than the facts and forecasts on future company financial performance, and such sentiment and attention might stem from social media in junction with traditional media (Baker & Wurgler, 2007; Barber & Odean, 2008; Bukovina, 2016; Kumar & Lee, 2006; Rakowski et al., 2021). Research by Luo et al. (2013) also demonstrates the impact on firm equity valuation by social media metrics such as " total blog posts, rating volume, total page views, and search intensity".

The difference in information obtained between traditional news media and online media can also be readily observed. The empirical researches of Brière and Ramelli (2021) and of Engle et al. (2020) suggest that the mentions of climate change may differ between printed media, Google search volume of related term and the published fund flow. Specifically, Engle et al. (2020) constructed an index of news sentiment from counting WSJ article mentions of environment issues using content analysis methods, and Brière and Ramelli (2021) noted the lack of correlation between Engle et al.'s

(2020) index and Google search volume.

ESG and Green Investment

Consideration of Environmental, Social and Governance (ESG) issues as factors for investment were first mentioned through the UN Global Compact initiative in 2004, in a report titled *Who Cares Wins*. The report pointed out the environmental, social and governance factors were only considered “if they are seen as being material to value creation and risk in the shortterm” (UN Global Compact Initiative, 2004). In 2007, a group of Swedish pension funds, with the goal to make a difference with investment, approached Sandinaviska Enskilda Banken AB, the Centre for International Climate and Environmental Research and the World Bank Treasury. With a collaborative effort, the world’s first green bond was created by the World Bank in November 2008 (World Bank, 2018). As noted by Antoncic (2019), the “asset under management by by signatories to the Principles for Responsible Investment” has risen to US$81 trillion in 2018. In a meta-analysis of researches regarding ESG and financial performance, Whelan et al. (2020) “found consistent positive correlations between ESG and corporate financial performance”, however “that ESG investing returns were generally indistinguishable from conventional investing returns”, meaning ESG investment is unable to generate additional alpha in comparison to traditional strategies. As the investment community increases its engagement in ESG and green investing, debates and claims continue to enshroud this topic.

With a multitude of commonly agreed sustainability report framework, a lack of agreement in ESG ratings among commercial rating agencies, and an infrequent ESG report period, it is difficult to reach a consensus on the sustainability measures of a company (Antoncic, 2019; Antoncic et al., 2020). Hence, agenda-setting might play a role in directing ESG investment allocations, using the aforementioned arguments by Pollach (2014) and Zucker (1978) on media acting as an intermediate layer on non-directly-observable topics. Even if some substantial and publicly available data are available on corporate sustainability issues, disgreements and bias can be introduced when selecting data and filling in missing data to make informed decisions (Kotsantonis & Serafeim, 2019). Therefore, the ESG investment topic is ripe to explore the potential of incorporating an agenda-setting perspective.

While ESG investment encompasses not only environmental aspect but also social and governance aspects, the majority of attention in the debate is paid to the environmental issues, and the environmental aspect indeed possesses more salient information. Hence, this research would concentrate on the environmental aspect of ESG.

Research Question

To test the hypothesis, the research intends to determine whether relationships exist between the social media posts regarding climate change and the market activities of stocks incidental to environmental issues. Following empirical researches on social media, stock market and climate sentiment, this research further examines (See Ardia et al., 2020; Santi, 2020; Schmidt, 2019):

(1) If including ex-post Tweet volume regarding climate change in addition to existing financial factors could better describe the monthly price returns of companies in alternative energy, oil and gas energy, and utilities sector that are publicly listed in the North America or have issued American Depositary Receipts (ADRs);

(2) If ex-post Tweet amount regarding climate change correlates with the green sentiment index from "monthly abnormal flow into environment-friendly ETFs", as proposed by Brière and Ramelli (2021);

(3) Whether ex-post Tweet volume correlates with news coverage by New York Times and Associated Press, similar to Engle et al.'s (2020) analysis of Wall Street Journal.

While question (1) and (2) directly concern the main hypothesis, question (3) would serve as an extention to the original hypothesis, in response to the discrepancy found by Brière and Ramelli (2021) between their green sentiment index and the index of WSJ climate change topic proposed by Engle et al. (2020). It would assist in analyzing whether the relationship is stronger between Tweet volume and the stock market, or between Tweet volume and written news intended for the consumption by the broader society.

Owing to the recency of the establishments of alternative energy companies and the data availability, the period of research data is set between April 2012 and April 2018.

Data

Tweets containing the phrases or hashtags "climate change", "global warming", "#climatechange", "#globalwarming" were retrieved from Twitter using Tweet status ID published by Samantray and Pin (2019) toward their research[2]. Since the research did not have access to full archive search of Twitter, an n-gram search of other climate change related terms in the Twitter history similar to Engle et al.'s (2020) method cannot

[2] Data available at https://doi.org/10.7910/DVN/LNNPVD

be conducted. Furthermore, Santi (2020) has demonstrated that introducing more search terms might introduce additional non-related noise into the social media data as well. The original file contains 14,353,859 Tweet status IDs without content between 2008 and 2018, from which 9,468,666 Tweets were able to be retrieved from Twitter using *rtweet* package in R (Kearney, 2019). By filtering the keywords to include only Tweets in English and filtering the post date to be between April 2012 and April 2016, 6,283,017 Tweets were made available for analyzation. Tweets between February 2012 and April 2012 were retained to study the time lag effect in agenda-setting theory. The deletion of historical Tweets would pose marginal impact to the research, assuming the amounts of Tweet deleted per month are only contingent on the total amounts of Tweet in the month, and the sentiment follows the same distribution of the total Tweets in the month. The change of available Tweet amount would follow the change of original Tweet amount.

For question (1), researches such as those conducted by Hsu et al. (2020) and Santi (2020) have used Standard Industrial Classification (SIC) and Industry Classification Benchmark (ICB) to identify clean and polluting companies, but the company query result is broad and inaccurate. For example, financial institutions would be classified as clean companies, whereas semiconductor companies that specialize in clean energy technology would be classified as polluting companies. Hence, to accurately capture the sector of companies, the research selected 23 constituent companies from NASDAQ Clean Edge Green Energy Index, S&P Global Clean Energy Index and S&P Kensho Cleantech Index as alternative energy companies, 52 companies from Dow Jones U.S. Oil & Gas Index as oil and gas energy companies and

46 companies from Dow Jones U.S. Utility Index as utilities companies. The selected companies were publicly traded in US exchanges from April 2012 to April 2018. Daily close prices and volumes of the stocks were accessed through Bloomberg Professional (Bloomberg L.P., n.d.). To minimize individual stock variation's influence on analysis result, four equal weighted portfolios were formed: alternative energy, oil and gas energy, utilities and green-minus-brown. The green-minus-brown portfolio is a long-short portfolio formed from buying alternative energy companies and selling oil and gas energy companies. The portfolios were formed and monthly portfolio arithmetic returns were calculated using tidyquant package from R.

Following Santi (2020) and Hsu et al. (2020), this paper assumes that the clean and polluting stock return is sensitive to Fama-French three factor model and the carbon price. Fama-French three factor model is a model for asset pricing built upon the capital asset pricing model (CAPM). The CAPM was developed by William Sharpe (1964), Jack Treynor (2002), John Lintner (1965, 1969) and Jan Mossin (1966) (Perold, 2004). The model takes risk into the acount of cost of capital, and incorporates market return into the modelling of individual stock return. The Fama-French three factor model introduces two additional factors to account for differences betwen stocks on the market: a high-minus-low factor captures the value premium of the spread between companies with a high book-to-market value ratio and companies with a low book-to-market value ratio; a small-minus-big factor captures the tendency of small companies outperforming big companies. Therefore, to account for these influncing factors, data for Fama-French three factor model was included as well. The Fama-French three factor model data is made available by French (n.d.), which is updated daily based on Fama and French's (1993) research result. The research selected the monthly Fama-French three factors data from the data library[3]. Santi (2020) noted that the carbon price might also be a contributing factor to the price of climate change related companies. Following Santi (2020) and to simply the model specification, this paper utilizes WTI Crude Oil spot price as a a proxy for spot carbon price. The monthly WTI Crude Oil spot price was obtained from U.S. Energy Information Administration (n.d.)[4]. The WTI Crude Oil spot price was transformed into the natural logarithmic of it.

For research question (2), to calculate the Green Sentiment Index, a total of 670 ETFs were indentified through applying Brière and Ramelli's (2021) criteria, namely ETF with equity long-short strategy and focused in North America. The fund total asset of 670 ETFs were retrieved from Bloomberg Professional and the total shares outstanding of these ETFs were retrieved from Refinitiv Datastream (Bloomberg L.P., n.d.; Refinitiv, n.d.). The net asset values (NAVs) were derived from these data as $NetAssetValue = \frac{FundTotalAsset}{TotalSharesOutstanding}$. Ten US-listed ETFs were marked as environment-friendly ETFs based on the list provided by Brière and Ramelli (2021). Following their research, the data is collected from January 2010 till the end of our research preiod of April 2018.

For research question (3), the research selected New York Times and Associated Press news data as comparision with Twitter data. The metadata of articles containing the keywords "global warming" and "climate change" in either the headline, the body or both were downloaded from LexisNexis (LexisNexis, n.d.). A total of 21,380 articles were recorded between April 2012 and April 2018. The aggregated amount can

[3] Dataset at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_Factors_CSV.zip
[4] See https://www.eia.gov/opendata/v1/qb.php?sdid=PET.RWTC.M

be seen in Figure 4.

Analysis

Climate Change Tweet Volume and Price Return of Stocks

To measure the sensitivity of stock price return to the Tweet volume, the research formed the four portfolios as mentioned above to reduce the likelihood of unsystemetic risk of each company affecting the result. Taking into account the Fama-French three factor model and the natural logarithm of WTI spot price, the following regressions were formed in describing the price return of a stock portfolio using monthly data:

$$R_{p,t} - R_{f,t} = \alpha + \beta_1 TweetCount_{t-k} + \beta_2(R_{m,t} - R_{f,t}) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 ln(WTISpot_t) + \epsilon_t \quad (1)$$

$$R_{p,t} - R_{f,t} = \alpha + \beta_1 TweetCount_{t-k} + \beta_2(R_{m,t} - R_{f,t}) + \beta_3 SMB_t + \beta_4 HML_t + \epsilon_t \quad (2)$$

$$R_{p,t} - R_{f,t} = \alpha + \beta_1(R_{m,t} - R_{f,t}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_t \quad (3)$$

where $k = \{0, 1, 2\}$ in attempt to compare the optimal time lag. $TweetCount_{t-k}$ is the total count of climate change related tweets in the month $t - k$. R_p is the return of the portfolio p at month t. Four factors from Fama-French three factor model are included in the regressions, where R_m is the return rate of market portfolio, R_f is the risk-free rate, SMB is small-minus-big factor and HML is high-minus-low factor (Fama & French, 1993). $WTISpot_t$ is the WTI spot price of month t. The risk-free rate is deteremined by one month U.S. Treasury Bill rate. Durbin-Watson Tests were performed to test for autocorrelation. In summary, equation (1) captures the sensitivity of excess return of a portfolio to the monthly Tweet amount, the Fama-French three factor model and the oil price; equation (2) captures the sensitivity of excess return of a portfolio to only the monthly Tweet amount and the Fama-French three factor model; equation (3) is the Fama-French three factor model to serve as a comparison with the first two regressions. The regression will yield the intercept and coefficients. If the agenda-setting effect of Twitter to the particular portfolio is prominent, the coefficient of $TweetCount_{t-k}$ will be statistically significant, and the model's adjusted R^2 score will be greater than that of the models without $TweetCount_{t-k}$. If the coefficient of $TweetCount_{t-k}$ given a particular time lag k_i is larger than the other coefficients with other time lags, then the time lag effect is more pronounced at k_i, and the effect exists for all ks where the coefficient is significant and greater than zero.

Furthermore, equation (3) is defined as the benchmark model to compare model fitting. The fitted model can better explain the data, if the adjusted R^2 score is higher than the benchmark model's adjusted R^2. The F-test was also introduced to test if the alternative hypothesis if the specified model is better than an intercept-only model. Since the data involves time series, the Durbin-Watson test was performed to test for autocorrelation. A high p-value in the DW test would signify presence of autocorrelation. The regression results are presented in the tables below.

Table 1

Regression Result of Climate Change Tweet Volume and Clean Portfolio Price Return

Portfolio	Clean						
Equation	(1)			(2)			(3)
k	0	1	2	0	1	2	
Intercept	-8.228	0.01349	-3.263	0.4386	1.115	0.8074	-0.2355
$TweetCount$	-1.421e-6	-1.422e-5	-8.962e-6	-7.323e-6	-1.505e-5	-1.191e-5	
$R_m - R_f$	1.514****	1.494****	1.519****	1.515****	1.495****	1.527****	1.5535****
SMB	0.6592*	0.5757	0.6649*	0.6542*	0.5698	0.6630*	0.6409*
HML	-0.3149	-0.2098	-0.2721	-0.3328	-0.2036	-0.2580	-0.3027
$ln(WTISpot_t)$	1.952	0.2468	0.9161				
Adjusted R^2	0.3473	0.3566	0.3513	0.3518	0.3660	0.3598	0.3568
df	67	67	67	68	68	68	69
F-statistic p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DW-statistic p-value	0.22	0.178	0.204	0.264	0.25	0.292	0.214

Note: **** p < 0.001, *** p < 0.01, ** p < 0.05, * p < 0.1

Table 2

Regression Result of Climate Change Tweet Volume and Oil and Gas Portfolio Price Return

Portfolio	Oil and Gas						
Equation	(1)			(2)			(3)
k	0	1	2	0	1	2	
Intercept	1.415	-5.558	1.352	-0.2176	-0.8663	-0.3548	-1.0913*
$TweetCount$	-1.060e-5	1.028e-6	-9.642e-6	-9.493e-6	-2.507e-6	-8.408e-6	
$R_m - R_f$	1.195****	1.232****	1.231****	1.195****	1.236****	1.227****	1.2456****
SMB	0.4583*	-0.4553*	0.4568*	0.4592*	0.4301*	0.4576*	0.4420*
HML	1.065****	1.097****	1.144****	1.068****	1.123****	1.138****	1.1069****
$ln(WTISpot_t)$	-0.3677	1.051	-0.3840				
Adjusted R^2	0.5356	0.5263	0.5335	0.5422	0.5312	0.5401	0.5372
df	67	67	67	68	68	68	69
F-statistic p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DW-statistic p-value	0.508	0.574	0.472	0.646	0.654	0.580	0.656

Note: **** p < 0.001, *** p < 0.01, ** p < 0.05, * p < 0.1

Table 3

Regression Result of Climate Change Tweet Volume and Utilities Portfolio Price Return

Portfolio	Utilities						
Equation	(1)			(2)			(3)
k	0	1	2	0	1	2	
Intercept	5.053	-3.942	3.241	0.7609	-0.02127	0.5408	0.4479
$TweetCount$	-6.324e-6	8.181e-6	-3.013e-6	-3.401e-6	5.226e-6	-1.061e-6	
$R_m - R_f$	0.2986*	0.3333**	0.3196**	0.2981*	0.3364**	0.3138**	0.3161**
SMB	-0.1449	-0.1029	-0.1479	-0.1425	-0.1240	-0.1467	-0.1486
HML	-0.2470	-0.2810	-0.2109	-0.2382	-0.2586	-0.2202	-0.2242
$ln(WTISpot_t)$	-0.9667	0.8784	-0.6077				
Adjusted R^2	0.0166	0.02077	0.009143	0.02561	0.03122	0.02179	0.03554
df	67	67	67	68	68	68	69
F-statistic p value	0.2990	0.2723	<0.0001	0.22	0.1896	0.243	0.1403
DW-statistic p-value	0.244	0.394	0.266	0.254	0.276	0.218	0.168

Note: **** p < 0.001, *** p < 0.01, ** p < 0.05, * p < 0.1

Table 4

Regression Result of Climate Change Tweet Volume and Green-Minus-Brown Portfolio Price Return

Portfolio	Green-Minus-Brown						
Equation	(1)			(2)			(3)
k	0	1	2	0	1	2	
Intercept	-3.064	4.103	-0.3121	0.2746	0.8725	0.5161	0.2764
$TweetCount$	2.293e-6	-9.075e-6	-2.138e-6	1.968e-8	-6.641e-6	-2.737e-6	
$R_m - R_f$	0.03913	0.01619	0.03168	0.03955	0.01365	0.03346	0.03944
SMB	0.1549	0.1044	0.1585	0.1530	0.1217	0.1581	0.1530
HML	-0.7280***	-0.6728***	-0.7275***	-0.7349***	-0.6912***	-0.7247***	-0.7349***
$ln(WTISpot_t)$	0.7519	-0.7237	0.1864				
Adjusted R^2	0.07892	0.09167	0.07873	0.08975	0.1028	0.09213	0.1029
df	67	67	67	68	68	68	69
F-statistic p-value	<0.1	<0.05	<0.1	<0.05	<0.05	<0.05	<0.05
DW-statistic p-value	0.784	0.822	0.764	0.704	0.604	0.594	0.620

Note: **** p < 0.001, *** p < 0.01, ** p < 0.05, * p < 0.1

For the clean stock portfolio, the adjusted R^2s failed to exhibit a significant improvement in equation (1) and (2) as compared with equation (3) at all time lags, and the regression coefficients of both $TweetCount_{t-k}$ and $ln(WTISpot_t)$ failed to demonstrate statistical significance. The F-statistics reported p-values below 0.0001, rejecting the null hypotheses that all coefficients are equal to zero. While it can be determined that Tweet count and oil price do not show significant statistical relationship with the price of clean portfolio, it is also interesting to observe that the null hypotheses of the intercept and HML coefficient of the Fama-French three factor model are equal to zero cannot be rejected at all conventional significance level, while the market return premium coefficient is significant at $p < 0.0001$.

A similar result can be seen in the oil and gas portfoilo, where only the Fama-French three factor model have the intercept and all coefficients being statistically significant at $p < 0.1$. However, such significance does not correspond to an increase in the the adjusted R^2 of the Fama-French three factor model.

The adjusted R^2s for the utilities portfolio and green-minus-brown portfolios are lower than those of clean portfolio and oil and gas portfolio. For the utilities portfolio, only the F-test of equation (1) with $k = 2$ successfully rejected the null hypothesis at all conventional significance levels.

As shown above, model fitness variations are demonstrated between portfolios from different sectors, with the oil and gas portfolio showing the strongest evidence for the proposed regressions and utilities and green-minus-brown portfolios showing the weakest evidence.

The evidence above rejects the hypothesis of research question (1), indicating the social media discussion on Twitter using the proposed keywords is not statistically significantly related to the performance of the stocks in the proposed sectors, either within the same monthly priod or with the lag of one or two months. However, it is worth noticing that the orignial Fama-French three factor model in equation (3) also did not perform well within all of the portfolios, where only the interecept and coefficients of the oil and gas portfolio are statistically significant.

Two potential interpretations on the methodology can be made. First, with the number of stock samples selected to form the portfolios is relatively small, as compared to the larger portfolios formed by Santi (2020) and Hsu et al. (2020). While all of the companies included in the portfolios are contingent on the climate change topic, the market capitalization of the companies are concentrated. For example, the oil and gas companies and utilities companies are concentrated in the mid- and large-cap tiers, while the clean companies are concentrated in the small- and medium-cap tiers. The small selection also gives rise to insufficient diversification of the portfolios, leaving the portfolio return susceptible to the impact of one particular company's price movement, especially if the sentiment surrounding such company is potentially influenced by news or social media.

Second, while the Fama-french three factor model achieved an excellent fitting for the oil and gas portfolio, and in comparison rejected the hypothesis of climate change Tweet count influencing the price return of such portfolios, such result is not evident in other portfolios. Because the comparison of the models with Tweet count is made against the fitting of Fama-French three factor models, the inability to achieve a satisfactory fitting by Fama-French three factor models in the clean, utilities and green-minus-brown portfolios cannot rule out the influence of climate change Tweets on their price returns.

Climate Change Tweet Volume and Green Sentiment Index

Engle et al. (2020) used unigram and bigram of climate change related keywords to perform a content analysis of Wall Street Journal articles and formed a daily climate change sentiment index based on term frequency-inverse document frequency (tf-idf). In *Green Sentiment, Stock Returns, and Corporate Behavior*, Brière and Ramelli (2021) proposed a green sentiment index by measuring the abnormal inflow to "environment-friendly" ETFs, and compared the index to Engle et al.'s (2020) result. Brière and

Ramelli (2021) demonstrated a significant difference between Engle et al.'s (2020) index and theirs. This section will test if a similar discrepancy exists between their proposed green sentiment index based on ETF inflow and the climate change tweet volume collected by this research.

The abnormal environment-friendly ETF inflow was calculated as follows (Brière & Ramelli, 2021):

$$Flows_{i,t} = \frac{SharesOutstanding_{i,t}}{SharesOutstanding_{i,t-1}} - 1 \tag{6}$$

where i is the ETF and t is the month. The creation or redemption of shares by authorized participants is indicated by the fund inflow/outflow of the ETF. Brière and Ramelli (2021) specified fund inflow/outflow to measure the green sentiment because of the structure of ETF as open-end fund. In an open-end fund, the share outstanding of the fund is unlimited, meaning the investors can create and redeem shares with the fund management as they buy and sell the shares, resulting in the change of shares outstanding. In the case of ETF, while the shares are traded between investors on exchanges, only authorized participants can create or redeem the shares. Because an ETF is a fund of a portfolio of stocks, when the ETF's market price deviates from its underlying net asset value, an arbitrage opportunity occurs where arbitrageurs can trade the ETF and its underlying stocks for economical profit. Therefore, the authorized participant will initiate creation or redemption of shares to capture this mispricing and shift the share price of the ETF. This action, in turn, would signify the action of the shareholders and is associated with the sentiment of them.

To determine the abnormal inflow/outflow, Brière and Ramelli (2021) proposed the following cross-sectional regression on each month's ETF flows:

$$Flows_{i,t} = c_t + \gamma_t \times GreenETF_{i,t} + \delta_t \times controls_{i,t} + \epsilon_{i,t}, \forall t \qquad (7)$$

where $GreenETF$ is a dummy variable to indicate whether the ETF is marked as green ETF. The standardized estimate γ as a time series is defined as the green sentiment index. Brière and Ramelli (2021) included several factors in $controls$. However, they have not specified in their paper the exact weight of each control factor (Brière & Ramelli, 2021). For simplicity, this research implements the past month ln(NAV) and past month return as $controls$.

Pearson's correlation is determined between the monthly volume of climate change Tweets and the green sentiment index. The resulting correlation is -0.1859 (p = 0.1103), hence the null hypothesis of correlation equal to zero is not rejected at any conventional significance levels. The discrepancy is anticipated from Brière and Ramelli's (2021) research, and the noted discrepancy between their measurement and Engle et al.'s (2020) measurement of negative climate news in Wall Street Journal. This difference prompts the discussion of whether the Tweet volume relates closer to the article count in traditional media, rather than that of the market sentiment observed from transactions.

Figure 2

Monthly Tweet Count on Climate Change as Filtered by Keywords

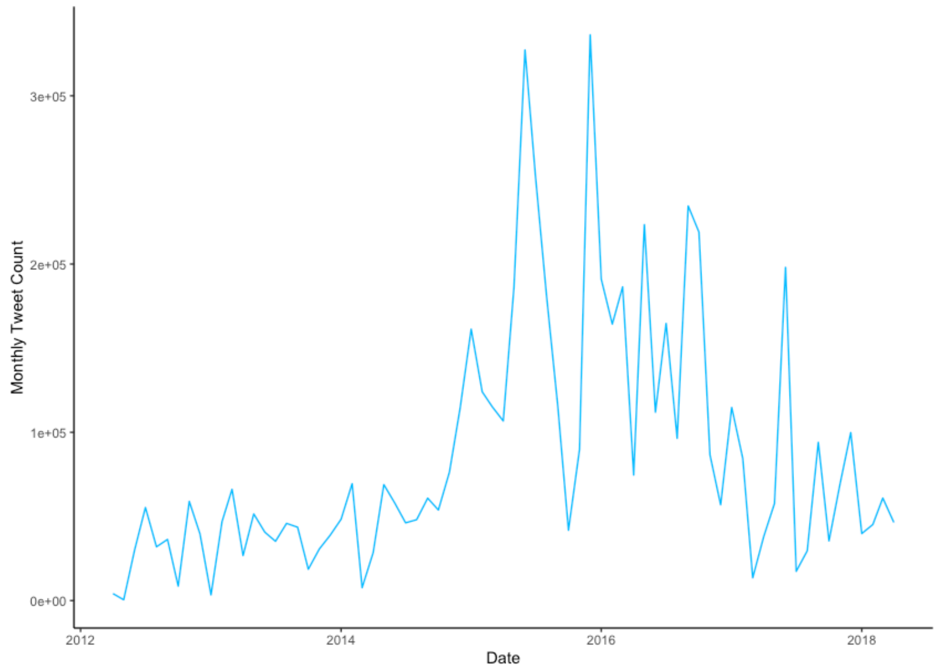


Figure 3

Result From a Modified Version of Green ETF Index as Proposed by Brière and Ramelli (2021)

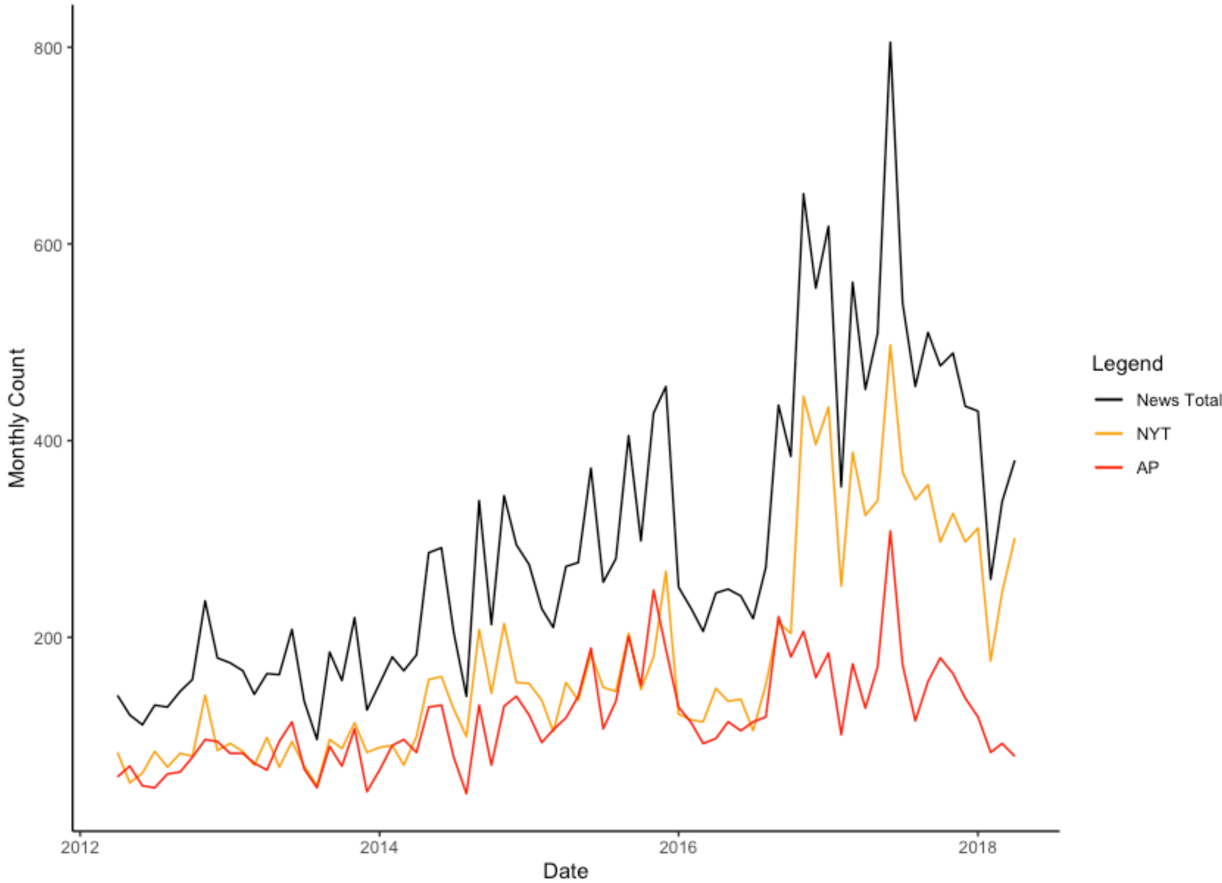
Climate Change Tweet Volume and News Volume

Due to the discrepancy in the previous section, and the relationship between news media and social media as noticed by Jiao et al. (2016) during their research, the paper proceeds to measure the relationship between news volume and Tweet volume on climate change. Engle et al. (2020) used a list of climate change terms transformed into unigrams and bigrams to measure the sentiment on WSJ. However, Santi (2020) noticed that including broader terms introduced query result on Stocktwits unrelated to climate change. In the climate change vocabulary list constucted by Engle et al. (2020), words such as "economic", "international", "period" and "refers" are considered relevant. Such word choice demonstrates the non-specificity of using these terms to filter climate change texts, even though the tf-idf score was utilized in their case to increase the weight of less commonly used terms. In addition, the Twitter data available to this paper is only limited to the ones as provided through Samantray and Pin's (2019) Tweet ID dataset containing limited keywords. Therefore, weighing the benefits and drawbacks of vocabulary selection methods and to set up a better comparison with less uncontrolled variable, the same search keywords used to filter the Twitter data except hashtags, "climate change" and "global warming", were used in the news query through LexisNexis.

In the given period, 17,523 and 21,536 articles fitting the criteria are published from Associated Press and New York Times, respectively. The individual and aggregated count data can be seen in Figure 4.

Figure 4

Monthly Counts of News Articles Gathered



To compare the relationship between Tweet volume and news articles, two correlations were considered and assumptions are made. The first correlation examines the relationship between the Tweet volume, and the article count from each agency and the combined total. This correlation assumes that the volume in each month is independent and does not exhibit autocorrelation with the volume relationship of the previous month.

The second correlation examines the residuals of Tweet volume, article count and the combined total, after fitting an AR(1) autogregressive model to these time series. The AR(1) model assumes the volume for each month is autocorrelated with the count in the previous month. The residuals from AR(1) model would indicate the excess change in Tweet volume and article count not predicted by data in the immediate previous period. The residuals can be seen in Figure 5. The tests yielded the results in Table 5 and 6.

Figure 5

Residuals of Monthly Counts After AR(1) Model Fitting

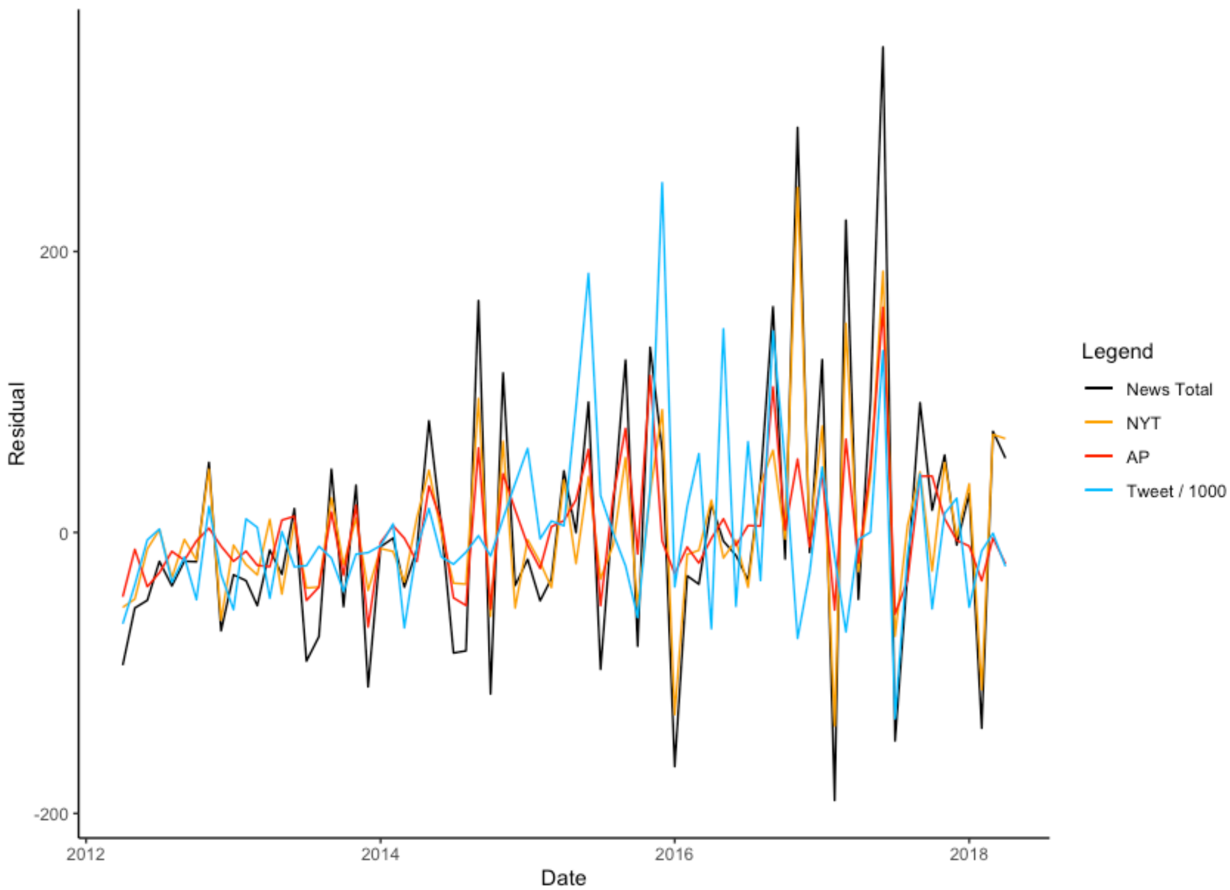


Table 5

Correlation between Tweet Volume and News Article Count

Publications	Pearson's Correlation
NYT	0.1463
AP	0.4860****
NYT and AP	0.2719**

Note: **** p < 0.001, *** p < 0.01, ** p < 0.05, * p < 0.1

Table 6

Correlation between the Residuals of Tweet Volume and News Article Count AR(1) Models

Publications	Pearson's Correlation
NYT	0.2324**
AP	0.3854****
NYT and AP	0.3029***

Note: **** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

By linearly regressing Tweet volume and news article count, regressing the residuals of Tweet volume and news article count, and applying Durbin-Watson test to both model specifications, the null hypotheses of no autocorrelation in Tweet volume and news article count time series fitted models are rejected ($p < 0.0001$), while the null hypotheses of no autocorrelation in the residuals fitted models are not rejected at all conventional significance levels ($p > 0.1$). Therefore, autocorrelation of period one exists in the relationships between Tweet volume and news article counts. The relationships between the Tweet volume and news article count residuals of AR(1) models are more robust.

Because the above models only account for data in each month, to examine whether time lag effects exist, the one and two month lagged values of Tweet volume residuals, or the one and two month lagged values of news article counts residuals were used to calculate correlation. The p-values of the correlations were found to be above 0.1. Time lag effect of equal or more than one month does not likely exist.

Discussion

In the analysis, the hypothesis posed by research question (1) and (2) are rejected, while the hypothesis in research (3) is statistically significant at 5%

significance level, after fitting AR(1) autoregressive models with the data and finding the correlations of residuals.

The following subsections will first discuss the significance of the result to each question, and then the overall implication on the relationship between the media and stock market.

Climate Change Tweet Volume and Price Return of Stocks

This research failed to identify the media agenda-setting effect on stock market. One possibility, as noted by Santi (2020), Brière and Ramelli (2021) and Bukovina (2016), is that the result stems from the difference in media consumption pattern of different investors. The retail investors are likely more susceptible to using social media information toward investment decisions, as they have less access to financial information and less restrictions placed on their trading behaviors, as explained by Zucker (1978) on media's influence on unobstrusive topics. On the other hand, institutional investors with more capital have more information available, are subject to more contractral responsibilities as well as legal scrutiny, hence most likely would resist trading on social media information.

The researches focused on idiosyncratic media metrics of specific companies have seen more significant relations among the social media, traditional media and stock market, which might be explained by agenda-setting effect, while the researches on the "stale" thematic information produced arbitrary measurements and yielded inconclusive effects between media and stock market (Brière & Ramelli, 2021; Engle et al., 2020; Jiao et al., 2016; Rakowski et al., 2021; Santi, 2020; Schmidt, 2019). This effect might be explained by the concept of rational inattention, where "attention is a scarce cognitive resource" and priorities must be given to the information most helpful to predicting uncertain outcomes (Kacperczyk et al., 2016; Kahneman, 1973; Peng & Xiong, 2006). While Peng and Xiong (2006) demonstrated that the investor allocates more attention to processing "market- and sector-level information" through "category learning", their research has also shown the investor can "allocate more attention" to "firm-specific information" when the attention constraint is less binding. They proposed that such belief on the information weight is incorporated "through Bayesian updating" that is based on previous result, which is in line with the hypothesized time lag in agenda-setting theory and the observation made by Jiao et al. (2016) (McCombs, 1981; McCombs et al., 1995; Peng & Xiong, 2006). Unfortunately due to the limited resource to acquire more granular and complete data from Twitter, or to construct an experiment in a controlled setting, this paper is unable to model such belief updating at intra-day or intra-month level with accuracy.

Linking the rational inattention concept to the case presented in this paper, the Tweet or news about climate change could be considered as sector-level information, and the financial factors from Fama-French three factor models could be considered as market-information. The overall cost of access to market-level and firm-specific information is much lower than sector-level information on climate impact, which is often plagued with disagreements, bias and missing data (Kotsantonis & Serafeim, 2019). This would lower the investor's expected return from sector-level information, including the deliberation occuring on news media and social media. Due to the lack of concrete actionable information on specific firm's stock, the news media and social media information would be marked as white noise and be omitted in the decision making process. Such postulation also corresponds to Herbert Simon's (1955) opinion on the rationality of economic agent as mentioned before. The goal of investors is to achieve the optimal rationality given decision making time and capacity, and social media post not relevant to immediate profiting opportunities must yield to other information in forming investment decisions. The overall agenda fits well within the adaptive market hypothesis, where each investor is concerned about her survival in the market and act according to their beliefs of best action (Lo, 2004).

Besides the adaptive market hypothesis and media consumption preference factors, several methodological limitations might also contribute to the rejection of hypothesis. As mentioned earlier in the result summary, the lack of diversification in the constructed portfolios might result in the insignificance of models. Considering the results from climate change tweet volume and price return of stocks section, the construction of the portfolios used in the analysis is different from those deployed by other researches (Brière & Ramelli, 2021; Engle et al., 2020; Jiao et al., 2016; Santi, 2020; Schmidt, 2019).

The lack of consensus in climate change vocabulary list is another contributing factor. The existing social media content analysis of climate change each proposed a different vocabulary list, and the classification result of climate change related content can vary significantly (See Engle et al., 2020; Samantray & Pin, 2019; Santi, 2020). A smaller selection of vocabulary can improve the precision of classification, but may come at the cost of increased type I error (climate change related content as null hypothesis). The constructed measurements may not capture the sentiment on social media in full.

The other methodological factor is the statistical modelling method used for research, which suggests the possible use of explainable artificial intelligence to fit both historical data explaination and prediction may yield more substantial results. All of the current researches on the topic have taken the data model approach rather than an algorithmic model approach, which would potentially achieve higher accurcy than data model (Breiman, 2001). The issue is evident in the lack of model fitting even by the Fama-French three factor model. A regression model is only capable of measuring single relationships between specified input factors and the output result. The potential interaction between the combinative result of factors cannot be modelled. With the hidden layers in a "black box" algorithmic model, or in a more specific term, deep neural network model, the model accuracy is increased at the trade off of observable causality (Breiman, 2001). In this case, the approach selection is highly contingent on the application.

Future potential research using neural network model must dealt with the model's inherent limitation. The data model approach is better suited toward historical data explaination, whereas the algorithmic model approach is better suited toward future trend prediction. While research using neural network might benefit from increased model fitness, such model cannot be used to explain historical trends. Nevertheless, there is also the possiblity of using explainable machine learning model to address such drawback. Carta et al. (2021) has proposed an explainable machine learning model in predicting stock market from news data using domain-specific lexicon and decision tree, but the model explaination is made regarding the threshold level used in the decision tree nodes, rather than the selection of nodes and its rationality. Hence, the meaning of the node can only be interpreted in the context of the model with low transferability to tasks outside of the model specification, such as the agenda-setting effect that this reserach seek to measure. This is still a potential direction for future researches to assume.

Climate Change Tweet Volume, Green Sentiment Index and News Volume

While the climate change Tweet volume differs from the green sentiment index proposed by Brière and Ramelli (2021), the correlation study between climate change Tweet volume and news article count yielded statistically significant outcome. The result might support the claim that agenda-setting effect exists between traditional media and social media regarding climate change, especially at a nationwide agency such as Associated Press rather than a local agency such as New York Times, but not among the two media sources and and financial signals. Because mental and financial cost of following traditional news source to post climate change Tweets is lower, the Twitter users are potentially more keen on following the trend in media than investors.

From the result of this research and the other similar researches, it could be argued that agenda-setting effect explains the increase in traditional media mentions simultaneously occuring with or leading the increase in Tweet volume, albeit the correlation is weaker in nature as opposed to the high correlation as demonstrated in Chapel Hill study (Brière & Ramelli, 2021; Engle et al., 2020; Jiao et al., 2016; McCombs & Shaw, 1972). Because the monthly data is the additive effect of weekly data, a time lag between news article count and Tweet volume might be possibly identified at sub-month level, by using daily or weekly data. It is also worth noticing that autocorrelation exists in the non-autoregressive model. This can indicate that the news volume of the current month is influenced by the volume of the past month.

Further Implications – Media and the Stock Market Beyond Climate Change

As the paper failed to identify media agenda-setting effect on stock market in the climate change topic, the media's hopeful role to unify environment protection and business cannot be identified (Walley & Whitehead, 1994).

Relating the discrepancy among this paper's result and indices constructed and compared by Brière and Ramelli (2021) and Engle et al. (2020), and the confirmation of social media's influence on idiosyncratic company's stock performance and volume by several researches, it gives rise to the debates on the mode of media as a contributor to market bubbles, and the appropriateness of using media content as financial signals (Kumar & Lee, 2006; Luo et al., 2013; Rakowski et al., 2021; Shiller, 2015). Succinctly summarizing the varied results between researches and accounts in literatures, social media is a probable contributor to the market pricing and trading activity under the following conditions:

1. The market operates under adaptive market hypothesis (Lo, 2004);
2. The combined cost of access to and analyzation of fundamental finance information is higher than that of media information, which is a condition ripe for media agenda-setting effect (McCombs & Shaw, 1972; Zucker, 1978);
3. Attention is limited and is prioritized toward firm-specific media content than thematic media content, and firm-specific media content is abundant (Kacperczyk et al., 2016; Kahneman, 1973; Peng & Xiong, 2006; Simon, 1955; Tumasjan et al., 2021);
4. Social media poses greater impact on market than traditional media, and such effect can be combined between social media and traditional media (Rakowski et al., 2021).

Additionally, it can also be hypothesized that agenda-setting effect exists between traditional news media and social media (Jiao et al., 2016).

Conclusion

To examine the agenda-setting effect from the social media to the stock market performance regarding climate change, the research examined the relationship between Tweet volume on climate change and sector-specific stock return, Tweet volume on climate change and abnormal fund inflow to green ETF, and Tweet volume and AP and NYT news count on climate change, from April 2012 to April 2018. After selecting and calculating the price return of four equal weight portfolios, namely clean, oil and gas, utilities and green-minus-brown, the research did not find statistically significant evidence of agenda-setting effect between the monthly Tweet volume and the returns of the portfolios with zero to two month period lag. The research also failed find relationship between the monthly Tweet volume and a modification of the green sentiment index of green ETF inflow, proposed by Brierè and Ramelli (2021). However, there exist statistically significant correlations between the residuals of AR(1) models of monthly Tweet volume and NYT and AP news articles, suggesting the potential existence of agenda-setting between news and social media as observed by Jiao et al. (2016).

The reserach proposes three potential explainations for the result between social media and equity market: the rational inattention by investors to exclude stale and non-firm-specific information on Twitter about climate change, the lack of a defined vocabulary list to gauge Twitter climate change volume, and the limitation of regression analysis in fitting results. Even through the research cannot prove the direct result from social media toward equity market, the alternative explaination of rational inattention of social media fits within the adaptive market hypothesis (Lo, 2004).

Regarding Twitter and traditional media on climate change topic, weak correlation of residuals of AR(1) model in support of weak media agenda effect from traditional media to social media is present. It would be interesting to observe whether Engle et al.'s (2020) climate change sentiment index constructed from Wall Street Journal can demonstrate similar correlation with the Twitter data as collected in this research. The current data cannot support the extended hypothesis that a time lag exists between the traditional media and social media on climate change topic, and no agenda-setting effect at or beyond the time lag of one month is identified. A dataset with more grainular and complete daily or weekly information would be required to research the time lag in agenda-setting below one month.

Generalizing beyond the scope of climate change topic with discussions from other researches and literatures, this paper argues that social media, in combination with traditional media, is a probable contributor to the market pricing and trading activity under four specified conditions, but not on a thematic level (Kacperczyk et al., 2016; Kahneman, 1973; Lo, 2004; McCombs & Shaw, 1972; Peng & Xiong, 2006; Rakowski et al., 2021; Simon, 1955; Tumasjan et al., 2021; Zucker, 1978).

It is difficult to fathom from this paper whether an increased media report volume on climate change will lead to an increased incentive for investors to divest from polluting industries and shift toward green investment, and if such unity between environment and business can only remain a wishful thinking (Walley & Whitehead, 1994). Yet, to end on a more positive note, it is confident to anticipate, that the more traditional media reports on climate change, the greater the citizen environment protection engagement in their daily lives and on social media (Ader, 1995; Olausson, 2011).

References

Ader, C. R. (1995). A Longitudinal Study of Agenda Setting for the Issue of Environmental Pollution. *Journalism & Mass Communication Quarterly*, *72*(2), 300–311. https://doi.org/10.1177/107769909507200204

Antoncic, M. (2019). Why Sustainability? Because Risk Evolves and Risk Management Should Too. *Journal of Risk Management in Financial Institutions*, *12*(3), 206–216.

Antoncic, M., Bekaert, G., Rothenberg, R. v, & Noguer, M. (2020). Sustainable Investment - Exploring the Linkage between Alpha, ESG, and SDG's. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3623459

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2020). Climate Change Concerns and the Performance of Green Versus Brown Stocks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3717722

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Beckett, K. (1994). Setting the Public Agenda: "Street Crime" and Drug Use in American Politics. *Social Problems*, *41*(3), 425–447. https://doi.org/10.2307/3096971

Bennett, W. (2003). Communicating Global Activism. *Information, Communication & Society*, *6*(2), 143–168. https://doi.org/10.1080/1369118032000093860a

Benton, M., & Frazier, P. J. (1976). The Agenda Setting Function of the Mass Media At Three Levels of “Information Holding.” *Communication Research*, *3*(3), 261–274. https://doi.org/10.1177/009365027600300302

Bloomberg L.P. (n.d.). *Bloomberg Professional*.

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3). https://doi.org/10.1214/ss/1009213726

Brière, M., & Ramelli, S. (2021). Green Sentiment, Stock Returns, and Corporate Behavior. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3850923

Bukovina, J. (2016). Social Media Big Data and Capital Markets — An Overview. *Journal of Behavioral and Experimental Finance*, *11*, 18–26. https://doi.org/10.1016/j.jbef.2016.06.002

Cahan, S. F., Chen, C., Chen, L., & Nguyen, N. H. (2015). Corporate Social Responsibility and Media Coverage. *Journal of Banking & Finance*, *59*, 409–422. https://doi.org/10.1016/j.jbankfin.2015.07.004

Carta, S. M., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021). Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting. *IEEE Access*, *9*, 30193–30205. https://doi.org/10.1109/ACCESS.2021.3059960

Castells, M. (2011a). *The Power of Identity*. John Wiley & Sons.

Castells, M. (2011b). *The Rise of the Network Society*. John Wiley & Sons.

Castells, M. (2013). *Communication Power*. Oxford University Press USA - OSO.

Cerf, V. (1993). How the Internet Came to be. *The On-Line User's Encyclopedia: Bulletin Boards and Beyond. Reading, Massachusetts: Addison-Wesley*.

Cohen, B. C. (1963). *Press and Foreign Policy*. Princeton University Press. https://doi.org/10.1515/9781400878611

Cook, F. L., Tyler, T. R., Goetz, E. G., Gordon, M. T., Protess, D., Leff, D. R., & Molotch, H. L. (1983). Media and Agenda Setting: Effects on the Public, Interest Group Leaders, Policy Makers, and Policy. *Public Opinion Quarterly*, *47*(1), 16. https://doi.org/10.1086/268764

Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature*, *36*(1), 47–74.

Elton, M. C. J., & Carey, J. (2013). *The Prehistory of the Internet and its Traces in the Present: Implications for Defining the Field* (W. H. Dutton, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199589074.013.0002

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News. *The Review of Financial Studies*, *33*(3), 1184–1216. https://doi.org/10.1093/rfs/hhz072

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fan, M., Stallaert, J., & Whinston, A. B. (2000). The Internet and the Future of Financial Markets. *Communications of the ACM*, *43*(11), 82–88. https://doi.org/10.1145/353360.353368

French, K. (n.d.). *Kenneth R. French – Data Library*. Retrieved June 27, 2022, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Ghanem, S. I. (1997). Filling the tapestry: The Second Level of Agenda-Setting. *Communication and Democracy*, 3–15.

Grossberg, S., & Gutowski, W. E. (1987). Neural Dynamics of Decision Making Under Risk: Affective Balance and Cognitive-Emotional Interactions. *Psychological Review*, *94*(3), 300–318. https://doi.org/10.1037/0033-295X.94.3.300

Haddock-Fraser, J. E., & Tourelle, M. (2010). Corporate Motivations for Environmental Sustainable Development: Exploring the Role of Consumers in Stakeholder Engagement. *Business Strategy and the Environment*, *19*(8), 527–542. https://doi.org/10.1002/bse.663

Hsu, P.-H., Li, K., & Tsou, C.-Y. (2020). The Pollution Premium. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3578215

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI*

Conference on Web and Social Media, *8*(1), 216–225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Iyengar, S., Peters, M. D., & Kinder, D. R. (1982). Experimental Demonstrations of the “Not-So-Minimal” Consequences of Television News Programs. *American Political Science Review*, *76*(4), 848–858. https://doi.org/10.2307/1962976

Jiao, P., Veiga, A., & Walther, A. (2016). Social Media, News Media and the Stock Market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2755933

Kacperczyk, M., van Nieuwerburgh, S., & Veldkamp, L. (2016). A Rational Theory of Mutual Funds’ Attention Allocation. *Econometrica*, *84*(2), 571–626. https://doi.org/10.3982/ECTA11412

Kahneman, D. (1973). *Attention and Effort* (Vol. 1063). Citeseer.

Kearney, M. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, *4*(42), 1829. https://doi.org/10.21105/joss.01829

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, *31*(2), 50–58. https://doi.org/10.1111/jacf.12346

Kumar, A., & Lee, C. M. C. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance*, *61*(5), 2451–2486. https://doi.org/10.1111/j.1540-6261.2006.01063.x

LexisNexis. (n.d.). *LexisNexis*.

Lins, K. v., Servaes, H., & Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *The Journal of Finance*, *72*(4), 1785–1824. https://doi.org/10.1111/jofi.12505

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance*, *20*(4), 587. https://doi.org/10.2307/2977249

Lintner, J. (1969). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets: A Reply. *The Review of Economics and Statistics*, *51*(2), 222. https://doi.org/10.2307/1926735

Lippmann, W. (1922). The World Outside and the Pictures in our Heads. In *Public Opinion* (pp. 3–32). MacMillan Co. https://doi.org/10.1037/14847-001

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, *30*(5), 15–29. https://doi.org/10.3905/jpm.2004.442611

Lo, A. W., & Repin, D. v. (2002). The Psychophysiology of Real-Time Financial Risk Processing. *Journal of Cognitive Neuroscience*, *14*(3), 323–339. https://doi.org/10.1162/089892902317361877

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *American Economic Review*, *90*(2), 426–432. https://doi.org/10.1257/aer.90.2.426

Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, *24*(1), 146–163. https://doi.org/10.1287/isre.1120.0462

Marill, T., & Roberts, L. G. (1966). Toward a Cooperative Network of Time-shared Computers. *AFIPS ’66 (Fall): Proceedings of the November 7-10, 1966, Fall Joint Computer Conference*, 425. https://doi.org/10.1145/1464291.1464336

McCombs, M. E. (1981). The Agenda-Setting Approach. *Handbook of Political Communication*, 121–140.

McCombs, M. E. (1992). Explorers and Surveyors: Expanding Strategies for Agenda-Setting Research. *Journalism Quarterly*, *69*(4), 813–824. https://doi.org/10.1177/107769909206900402

McCombs, M. E. (2005). A Look at Agenda-setting: Past, Present and Future. *Journalism Studies*, *6*(4), 543–557. https://doi.org/10.1080/14616700500250438

McCombs, M. E., Danielian, L., & Wanta, W. (1995). Issues in the News and the Public Agenda: The Agenda-Setting Tradition. *Public Opinion and the Communication of Consent*, 281–300.

McCombs, M. E., & Ghanem, S. I. (2001). The Convergence of Agenda Setting and Framing. In S. D. Reese, Jr. Gandy, & A. E. Grant (Eds.), *Framing Public Life: Perspectives on Media and Our Understanding of the Social World* (pp. 67–81). Routledge. https://doi.org/10.4324/9781410605689

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, *36*(2), 176. https://doi.org/10.1086/267990

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate Social Responsibility and Firm Financial Performance. *Academy of Management Journal*, *31*(4), 854–872. https://doi.org/10.2307/256342

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, *34*(4), 768. https://doi.org/10.2307/1910098

Olausson, U. (2011). “We’re the Ones to Blame”: Citizens’ Representations of Climate Change and the Role of the Media. *Environmental Communication*, *5*(3), 281–299. https://doi.org/10.1080/17524032.2011.585026

Patterson, T. E. (1993). *Out of Order*. A. Knopf.

Peng, L., & Xiong, W. (2006). Investor Attention, Overconfidence and Category Learning. *Journal of Financial Economics*, *80*(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Perold, A. F. (2004). The Capital Asset Pricing Model. *Journal of Economic Perspectives*, *18*(3), 3–24. https://doi.org/10.1257/0895330042162340

Peters, E., & Slovic, P. (2000). The Springs of Action: Affective and Analytical Information Processing in Choice. *Personality and Social Psychology Bulletin*, *26*(12), 1465–1475. https://doi.org/10.1177/01461672002612002

Pollach, I. (2014). Corporate Environmental Reporting and News Coverage of Environmental Issues: an Agenda-Setting Perspective. *Business Strategy and the Environment*, *23*(5), 349–360. https://doi.org/10.1002/bse.1792

Rakowski, D., Shirley, S. E., & Stark, J. R. (2021). Twitter Activity, Investor Attention, and the Diffusion of Information. *Financial Management*, *50*(1), 3–46. https://doi.org/10.1111/fima.12307

Refinitiv. (n.d.). *Refinitiv Datastream*.

Robbins, S., & Judge, T. (2016). *Organizational Behavior, EBook, Global Edition*. Pearson Education, Limited. http://ebookcentral.proquest.com/lib/londonschoolecons/detail.action?docID=5185702

Robinson, D., & Coar, K. (2004). *The Common Gateway Interface (CGI) Version 1.1*. https://doi.org/10.17487/rfc3875

Samantray, A., & Pin, P. (2019). Credibility of climate change denial in social media. *Palgrave Communications*, *5*(1), 127. https://doi.org/10.1057/s41599-019-0344-4

Samuelson, P. A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, *6*, 41–49.

Sandvig, C. (2013). *The Internet as Infrastructure* (W. H. Dutton, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199589074.013.0005

Santi, C. (2020). Investors' Climate Sentiment and Financial Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3697581

Schmidt, A. (2019). Sustainable News – A Sentiment Analysis of the Effect of ESG Information on Stock Prices. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3809657

Schneider, S. H. (1989). *Global Warming: Are We Entering the Greenhouse Century*. San Francisco, CA (USA); Sierra Club Books. https://doi.org/https://doi.org/

Sharpe, W. F. (1964). Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Shiller, R. J. (2015). *Irrational Exuberance*. Princeton University Press. https://doi.org/10.2307/j.ctt1287kz5

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, *43*(3), 377–391. https://doi.org/10.1177/00027649921955326

Treynor, J. L., & French, C. W. (2002). Jack Treynor's "Toward a Theory of Market Value of Risky Assets." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.628187

Tumasjan, A., Braun, R., & Stolz, B. (2021). Twitter Sentiment as a Weak Signal in Venture Capital Financing. *Journal of Business Venturing*, *36*(2), 106062. https://doi.org/10.1016/j.jbusvent.2020.106062

UN Global Compact Initiative. (2004). *Who Cares Wins: Connecting Financial Markets to a Changing World*.

Ungar, S. (1992). The Rise and (Relative) Decline of Global Warming as a Social Problem. *The Sociological Quarterly*, *33*(4), 483–501. https://doi.org/10.1111/j.1533-8525.1992.tb00139.x

U.S. Energy Information Administration. (n.d.). *Cushing, OK WTI Spot Price FOB, Monthly*. Retrieved June 27, 2022, from https://www.eia.gov/opendata/v1/qb.php?sdid=PET.RWTC.M

Walley, N., & Whitehead, B. (1994, May). It’s Not Easy Being Green. *Harvard Business Review*.

Wanta, W., & Ghanem, S. (2007). Effects of Agenda Setting. In *Mass Media Effects Research: Advances Through Meta-Analysis* (pp. 37–51). Lawrence Erlbaum Associates Publishers.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community. *Annual Review of Sociology*, *22*(1), 213–238. https://doi.org/10.1146/annurev.soc.22.1.213

Whelan, T., Atz, U., van Holt, T., & Clark, C. (2020). *ESG and Financial Performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 – 2020*. https://www.stern.nyu.edu/sites/default/files/assets/documents/ESG%20Paper%20Aug%202021.pdf

Winter, J. P., & Eyal, C. H. (1981). Agenda Setting for the Civil Right Issue. *Public Opinion Quarterly*, *45*(3), 376. https://doi.org/10.1086/268671

World Bank. (2018, November 27). *From Evolution to Revolution: 10 Years of Green Bonds*. https://www.worldbank.org/en/news/feature/2018/11/27/from-evolution-to-revolution-10-years-of-green-bonds

Zucker, H. G. (1978). The Variable Nature of News Media Influence. *Annals of the International Communication Association*, *2*(1), 225–240. https://doi.org/10.1080/23808985.1978.11923728