Research Paper 70 | **2021**

HEARD THE NEWS? ENVIRONMENTAL POLICY AND CLEAN INVESTMENTS

Joëlle NOAILLY, Laura NOWZOHOUR, Matthias VAN DEN HEUVEL

Heard the News? Environmental Policy and Clean Investments

Joëlle Noailly[a,b,c*] Laura Nowzohour[a]
Matthias van den Heuvel[d]

[a]Graduate Institute Geneva, Switzerland
[b]Vrije Universiteit Amsterdam, The Netherlands
[c]Tinbergen Institute, The Netherlands
[d]EPFL, Switzerland

Abstract

We build a novel news-based index of US environmental and climate policy and examine how it relates to clean investments. Extracting text from ten leading US newspapers over the last four decades, we use text-mining techniques to develop a granular news index of US environmental and climate policy (EnvP) over the 1981-2019 period. Furthermore, we develop a set of additional measures, namely an index of sentiment on environmental policy, as well as various topic-specific indices. We validate our index by showing that it correctly captures trends and peaks in the evolution of US environmental and climate policy and that it has a meaningful association with clean investments, in line with environmental regulations supporting growing opportunities for clean markets. In firm-level estimations, we find that our index is associated with a greater probability of receiving venture capital (VC) funding for cleantech startups and reduced stock returns for high-emissions firms most exposed to environmental regulations. At the aggregate level, we find in VAR models that a shock in our news-based index of renewable energy policy is associated with an increase in the number of clean energy VC deals and in the assets under management of the benchmark clean energy exchange-traded fund.

Keywords: Environmental policy, News and media, Text-mining, Machine learning, Clean technologies, Investments, Venture Capital, Stock Returns.
JEL-Classification: Q58; C55; D81; E22

*Corresponding author: Joëlle Noailly, joelle.noailly@graduateinstitute.ch, Tel: +41229086222. Fax: +41229084225, Address: Center for International Environmental Studies and Department of International Economics, Graduate Institute of International and Development Studies, Chemin Eugene-Rigot 2, 1202 Geneva, Switzerland. This study is supported by the Swiss National Science Foundation (SNSF) within the framework of the National Research Programme "Sustainable Economy: resource-friendly, future-oriented, innovative" (NRP 73, Grant-No 407340-172395). We thank Luc Mottin, Julien Gobeill (Swiss Institute of Bioinformatics), Eva Bortolotti, Bettina Kast, Dario Siegen, Sarah Schoch, Mihnea Filip and Till Fust (Graduate Institute Geneva) for excellent research assistance. We thank Patrick Ruch, Gaetan de Rassenfosse, Raphael Calel, Nathan Sussman, Cédric Tille, Stefano Ramelli, Lorenzo Bretscher, Shuang Chen, Ioannis Branikos, Paolo Crosetto and participants at the 2020-2021 NBER Workshops on the Economics of Energy Innovation, SFI Research Days, the 2021 EEA Congress, FAERE and EAERE conferences, and the IMF's climate research seminar for helpful comments.

1 Introduction

This article presents the first news-based index of US environmental and climate policy, available on a monthly basis over the 1981-2019 period. We apply text-mining techniques to articles from ten leading US newspapers to construct a general news-index capturing the history of US federal and state-level environmental and climate regulations. We also develop a set of additional measures, namely an index of sentiment on US environmental and climate policy and 25 topic-specific indices, such as renewable energy policy and international climate negotiations among others.[1]

We evaluate our general environmental and climate policy (EnvP) index by showing that it captures significant policy events in the history of US environmental regulations and that it co-moves with the evolution of environmental policy stringency. We further validate the index by verifying that it has a meaningful association with financial investments most exposed to environmental regulations. More specifically, we find in firm-level estimations that our EnvP index is associated with a greater probability of cleantech startups receiving VC funding and reduced stock returns for high-emissions firms most exposed to environmental policy. Moreover, we find in VAR models that a shock in our news-based index is associated with an increase in the number of clean energy deals at the macro level and an increase in the assets under management of the main clean energy exchange-traded fund, aligned with evidence that environmental policy fosters growing opportunities for clean markets. The EnvP index is available online here: https://www.financingcleantech.com/envp-index and can be freely accessed by researchers.

Our news-based measure of environmental policy provides several complementary insights to existing quantitative indicators of environmental regulations. First, because news arrive daily and are available over long periods of time, our news-based index is a significant improvement over existing indicators of environmental and climate policy computed on an annual basis. Newspapers provide a large volume of (relatively low-cost and easily accessible) information.[2] Our approach provides a continuous tracking of environmental policy over time at a high frequency (monthly and quarterly time series), making it possible to measure immediate

[1]Our index mostly focuses on regulations of environmental pollution (e.g. greenhouse gas emissions and other air pollutants from electricity generation, vehicles and buildings, water pollution, oil spills, toxic and hazardous waste, nuclear waste) and abstracts from policies regulating natural resources (e.g. on forests, groundwater extraction, fishery, etc).

[2]While we could argue that professional investors may rely on more sophisticated information channels, such as business news or social media, newspapers present the advantage of being available for much longer time periods and to be relatively low-cost compared to other media. In addition, we consider newspapers, such as *The Wall Street Journal* or *The New York Times*, which target the audience of investors.

market reactions and to better address unobserved heterogeneity in empirical work (Ghanem & Smith, 2020). By contrast to event studies, our index is able to capture long-lasting dynamics of the policy process (e.g. announcements, delays, revisions) and how these are associated with economic impacts on markets. Extracting information from newspapers may also provide additional information on the regulatory context – such as details on implementation, controversies and sentiment – which are not typically captured by standard indicators.

Second, our index is available at various levels of aggregation (generic or topic-specific) helping to address the challenge of the multidimensionality of environmental policy. Regulations are often introduced as a 'package' of policies covering multifaceted aspects (such as the Green New Deal) and governments typically regulate the polluting activities of households and industries across many sectors on a wide range of pollutants. Summarizing and quantifying these various aspects into meaningful composite indicators is a very challenging task in empirical work (Brunel & Levinson, 2016). By using machine learning techniques on the rich amount of text provided by news articles, we are able to disentangle (latent) information on various subclusters of environmental policy issues and to build topic-specific indices, tracking for instance the subsample of news on renewable energy policy or international climate negotiations over time.

Our work raises several concerns about significance, accuracy and potential bias which we evaluate in several ways. A first concern relates to what our index actually measures and how it relates to policy stringency. As environmental policy becomes more stringent, economic agents will have greater (lower) incentives to invest into clean (dirty) markets. In a similar fashion, an implicit assumption in our work is that an increase in the volume of environmental and climate policy news raises the awareness of economic agents about existing regulations and growing opportunities in clean markets, leading them to increase their clean investments. Even though investors receive a 'noisy' signal of the underlying state of US environmental and climate regulations via news, they can nonetheless form expectations based on the information received.[3] We first verify that environmental and climate policy news correlate with the evolution of regulatory stringency over time. We further validate our news-based index by showing that it is associated with financial investments – as proxied by venture capital funds and stock returns –

[3] Our index is noisy in the sense that it reflects both the state of environmental and climate policy (e.g. in terms of government costs, stringency, etc) and the intensity of media coverage on these policies. We do not attempt to disentangle the impact of regulation from the impact of media coverage – a highly challenging task that we leave for future research.

in a manner that is consistent with environmental regulations opening up growing opportunities in clean markets. An immediate related concern when focusing on the relative volume of news is that our index may inaccurately capture negative discussions about environmental policy – e.g. relating to the high costs of environmental regulations leading to opposition by lobbyists or regulatory rollbacks – giving rise to perceptions of a decline in stringency. We deal with this by showing that our results remain robust when adding controls for a measure of environmental policy sentiment.

Additional concerns relate to the accuracy of our index. Environmental problems (and their semantics) evolve over time and we may be worried about missing out important policies. We address this issue by relying on machine learning approaches, rather than manual labelling or refined keyword searches. These techniques present the advantage of easing the processing of large amounts of text and of uncovering latent patterns without imposing too much structure on the text – although we invariably provide a minimum level of critical data to train the algorithms and inform the models. Machine learning approaches also make it possible to quantify measurement errors and to assess the performance of various algorithms. We find that our supervised learning model predicts environmental policy articles and captures relative trends over time relatively well.[4]

An additional important concern is that our index may be affected by media bias. The production of news is the result of an equilibrium between demand- and supply-side factors. On the demand side, newspapers feed news according to the preferences of their readers towards specific topics (Gentzkow & Shapiro, 2010; Mullainathan & Shleifer, 2005). As readers become more aware about environmental problems, they may demand more reporting about these and the policies to address them. On the supply side, economic incentives, competitive pressures between topics, journalists' norms[5], and communication from policy organizations[6] all affect which topics are being covered on a daily basis (Baron, 2006). Competition between media outlets and readers' heterogeneity, however, tend to provide incentives to increase accuracy in the reporting of news and contribute to mitigate media bias (Mullainathan & Shleifer, 2005;

[4]Due to inevitable measurement errors, our index does not pretend to retrieve the whole universe of environmental policy articles. Trends are however correctly identified as long as the distribution of environmental policy news remain constant over time.

[5]There is for instance evidence that in the US journalists' ideological preferences and norms gave a misrepresentation of the scientific consensus on climate change (Boykoff & Boykoff, 2007; Brüggemann & Engesser, 2017; Shapiro, 2016), which could slant reporting about environmental policy in a negative way.

[6]Muehlenbachs et al. (2011) find that controversial press releases from the Environmental Protection Agency on enforcement actions or regulatory changes were issued more often on Fridays, a time when news has the least impact on media coverage.

Gentzkow & Shapiro, 2006). Accordingly, we include a wide range of newspapers in our analysis and investigate potential bias due to partisan readerships. As macroeconomic and political factors also influence the reporting of environmental policy news – for instance public support and interest for environmental policy typically falls during recessions (Kahn & Kotchen, 2011), we control for business and political cycles via time fixed effects in our empirical estimations.

Finally, we validate the fact that despite being a noisy signal of environmental and climate policy our news-based index has a meaningful positive association with clean investments. There are important potential endogeneity issues when looking at the relationship between our EnvP index and clean investments. First, journalists may be reporting relatively more about environmental and climate policy when clean markets are growing, for instance under the influence of green lobby groups. Second, omitted variables such as technological change or evolving consumer preferences may also affect both media coverage and market outcomes. As such, establishing causality in the absence of natural experiments is highly challenging. To make progress, nonetheless, we estimate whether our news-based index has a differentiated impact on investments in venture capital or stock returns of firms most exposed to environmental policy, as defined by sector of activity or emission levels. In (micro) firm-level estimations, we find a meaningful association between our index and clean investments. We further test the robustness of our results to many fixed effects specifications in order to reduce the influence of confounding factors. At last, we investigate the dynamic relationship between our index and aggregate clean investments in (macro) VAR settings. OVerall, we find that the positive association between our EnvP index and the growth of clean markets persists across various measures of clean investments, a large range of robustness specifications, and at both micro and macro levels.

Our paper innovates by being the first study to construct a news-based index of environmental and climate policy using machine learning techniques. There is a growing number of studies showing that news-based indices provide meaningful economic information, but no other work so far has extracted information from news to measure environmental and climate regulations. In macroeconomics, Baker et al. (2016) introduced the methodology to build indicators of economic policy uncertainty searching for keywords into news articles. Manela and Moreira (2017) rely on front-page articles of *The Wall Street Journal* to build a text-based measure of uncertainty using machine learning techniques to predict the co-movement between news data and implied volatility indices. They also use content analysis to highlight the importance of various topics, such as wars and government policy, into risk premia variations. Also related to our work, Bybee et al. (2020) conduct a topic model analysis of business news published in *The Wall Street Journal* over 1984-2017 and find that specific news topics – for instance on recessions – have a significant predictive power on future output and employment.

In contrast with developments in macroeconomics, applications of textual analysis of news and media in environmental economics remain limited (Dugoua, Dumas, & Noailly, 2022; Baylis, 2020), with the exception of a recent literature on finance and climate change using text-as-data methods to quantify climate risks (Sautner et al., 2020; Kölbel et al., 2020; Engle et al., 2020). Close to our work, Engle et al. (2020) collect climate change news from *The Wall Street Journal* to provide a measure of climate risks as perceived by investors. While their news-based index relates solely to the broad concept of climate change, our EnvP index refers instead explicitly to the regulatory and policy framework underlying a broad range of environmental concerns (including other air pollutants beyond greenhouse gas emissions, water pollution, oil spills, toxic and hazardous waste, etc). Because we measure US environmental and climate policy in a more precise way, we find that the association between our EnvP index and clean investments remains meaningful even after controlling for the Climate Change News index of Engle et al. (2020).[7]

Our paper also fits within the broad literature examining how environmental and climate policy affects economic outcomes (Cohen & Tubb, 2018; Greenstone, 2002). Existing studies looking at the impact of environmental regulations on clean markets mostly relies on event studies around implementation dates of specific policies (Kruse, Mohnen, Pope, & Sato, 2020; Kruse, Mohnen, & Sato, 2020; Mukanjari & Sterner, 2018; Barnett, 2019). Kruse, Mohnen, Pope, and Sato (2020) find that the stock returns of US firms developing green goods increased by 10% in the week following the Paris Agreement. Related work by Mukanjari and Sterner (2018) finds no evidence that neither the Paris Agreement nor the election of Trump in 2016 significantly affected the returns for fossil fuel stocks. While most of this work focuses on single policy events often defined in an artificially narrow time window, our index is unique in its ability to track the evolution and unfolding of policies and regulations over long periods of time. In addition, a major challenge in the environmental economics literature is to capture the

[7]Beyond economics and finance, there is a growing body of work in communication sciences which investigates climate change news by applying automated textual analysis and topic modeling of newspapers (Bohr, 2020; Keller et al., 2020; Dahal et al., 2019), but so far none of this work has specifically looked at environmental and climate *policies* as we do.

multidimensional aspect of environmental and climate regulations (Brunel & Levinson, 2016). Typically, environmental regulations are very complex and cover many pollutants across various sources and targets. Due to measurement issues, the literature has either focused on narrow environmental problems, used broad proxies to capture the intensity of regulations – such as grams of lead-content per gallon of gasoline as in Cole and Fredriksson (2009) or perceptions through surveys (Johnstone et al., 2012), or constructed composite indicators based on counts of environmental policy measures – such as the environmental policy stringency (EPS) index as in Albrizio et al. (2017) and Botta and Kozluk (2014). Our index adds to this literature by illustrating a novel way to decompose environmental and climate news into specific subtopics. More broadly, our contribution to the environmental economics literature is to bring in new data to quantify fine-grained and hard-to-measure aspects of US environmental and climate policy better fitting the needs of researchers.

The paper is organized as follows. Section 2 describes the data and methodology used to construct our news-based index of US environmental policy, as well as various descriptives and validity checks. Section 3 presents additional measures, namely a general measure of sentiment and various topic-specific indices, which can be directly derived from our general index. Section 4 examines the relationship between our environmental policy index and proxies for investments in clean markets, namely cleantech venture capital deals and stock returns – both at the firm-level in panel regressions and the aggregate level in VAR settings. Section 5 concludes.

2 A News-based Index of Environmental Policy

2.1 Developing a news-based index of environmental policy

Our analysis starts from a set of 15 million articles from the archives of ten US newspapers available via automated access through Dow Jones' Factiva platform over the 1981-2019 period. We obtain access to the following newspapers: *New York Times, Washington Post, Wall Street Journal, Houston Chronicle, Dallas Morning News, San Francisco Chronicle, Boston Herald, Tampa Bay Times, San Jose Mercury News and San Diego Union Tribune.* Table A1 in Appendix A provides detailed starting dates of the archive and additional statistics on the distribution of articles across newspapers.

As news about environmental regulations are relatively rare, manual retrieval of environmental policy articles is highly challenging. Hence, we first reduce our initial sample by selecting articles that contain keywords related to both 'climate change and the environment' and 'policy and regulations' within a proximity of 40 characters. Our choice of keywords in the category 'climate change and the environment' includes terms related to clean technologies[8], which cover energy generation, energy storage, energy efficiency, lighting, pollution (air, water, land), transportation (batteries, electric vehicles, clean fuels, etc.), recycling and waste and have been cross-checked against a broad definition of climate and environmental keywords from Climate Tagger.[9] The set of keywords on policy and regulations is borrowed from Baker et al. (2016) to which we add specific terms related to environmental policy (e.g. 'feed-in tariff').[10]

At this stage, our search strategy is extremely broad, as we want to avoid missing out on potentially relevant articles (type-II error) by imposing too much structure. We obtain a set of 459,000 articles resulting from our query. For now, a large set of these articles are irrelevant as many general terms such as 'environment' or 'climate' are used in a myriad of contexts not applicable to environmental and climate regulations (type-I error). These articles refer for instance to 'tax policies to improve the business climate', 'agreement for changing the political environment', 'sustainable plans', etc.

Training sets, text pre-processing and classification

Our objective is then to correctly identify articles on environmental policy within our set of 459,000 articles. To do so, we use supervised machine learning which has two attractive features. First, it circumvents the need for a manual labeling of the entire corpus. Second, machine learning imposes only a minimum level of structure on what constitutes a relevant article (i.e., an article about environmental and climate policy), in contrast with the more restrictive use of a combination of keywords.[11]

[8] We used the website `www.cleantech.org` for our main definition of clean technologies. This definition excludes topics of conservation, fisheries, forestry and biodiversity and natural resources. These issues are a priori less relevant for clean investments and markets.

[9] www.climatetagger.net

[10] The full set of keywords and Boolean operators are available upon request.

[11] Manela and Moreira (2017) propose a related methodology relying on machine learning to build a news-based index of macroeconomic uncertainty. They train a support vector based learning model by relating text from titles and abstracts from the *The Wall Street Journal* articles to an index of implied option volatility (VIX). By contrast to their work, we use manually labeled articles as a training set for our support vector algorithm. In addition, we consider the full length of news articles as well as a much larger set of newspapers. Our topic analysis in Section 3.2 also presents some similarity with Manela and Moreira (2017). Yet, while they use WordNet to categorize words into topics, we apply LDA topic modeling. A main difference is that WordNet uses a dictionary-based algorithm to group semantically similar texts into topics, while LDA examines the co-occurrence of words within a document to uncover topics. Both approaches lead to similar insights (and can be combined in more advanced methods), although there is some evidence that WordNet methods may be more suited for shorter texts and LDA for longer texts, justifying our preference for LDA in our case (Chen et al., 2012).

The first step when using a supervised learning approach is to build manually annotated training sets which inform the algorithm about the content of environmental policy articles. We start by reading a large number of newspaper articles in order to develop a codebook defining criteria to classify environmental policy articles.[12] We then randomly select sets of articles to build two training sets: 1) an initial set of 995 articles from the *New York Times*, due to its high editorial quality and because its archive could be crawled early on in the process, 2) a set of 1,469 articles from our whole sample of newspapers, which better reflects the diversity of environmental policy articles. Three annotators then separately review overlapping sets of articles and manually assess whether or not a given article discusses environmental policy, guided by our codebook classification. About 20% of articles in the training set are classified as relevant for environmental policy by the annotators.

As a second step, we apply a set of standard text pre-processing steps to our corpus of 459,000 articles (removing very short articles, removing html tags, numbers and punctuation, lowercasing all words, stop-words filtering and lemmatization). Following standard approaches in computational linguistics, we convert articles into numerical vectors of unigram and bigram frequencies using a 'bag-of-words' approach, i.e. disregarding grammar and word order.[13] We then construct a standard term-frequency inverse-document frequency (tf-idf) matrix in which less weight is given to words that occur either very frequently in the corpus or are barely used in the articles where they appear, because these are less informative than other words.[14]

In a third step, we use our training sets and the tf-idf matrix as inputs for a support vector machine (SVM). SVM is a predictive data-classification algorithm which learns from the training set how to assign labels (i.e., environmental and climate policy or not) to articles based on their most distinctive text features. We provide further details on the SVM algorithm, its parametrization and cross-validation in Appendix B.

We find that when the SVM model classifies an article as pertaining to environmental policy,

[12]Our full codebook is available on the website dedicated to our index: `https://www.financingcleantech.com/envp-index`

[13]Besides bag-of-words approaches, other more sophisticated models of deep learning rely on word vectors (word embedding). We prefer the bags-of-words approach, where every dimension in the vector space stands for a word or bigram, because it allows us to provide better interpretation of the decision rule of the classifier based on word features as in Table 1. More complex deep learning methods are more of a 'black box' with low transparency and interpretability as vector representations are harder to understand and therefore to validate by social scientists. We also see an added value in bags-of-words approach in helping researchers to build lexicons for future applications.

[14]More specifically, given a term-frequency matrix $tf(n,m)$, such that n is the number of articles and m the number of words, each term-frequency count is multiplied by the inverse document frequency. The Inverse document frequency ($idf_{j,n}$) is given by $log\left(\frac{N}{n_j}\right)$, where N is the total number of documents and n_j is the total number of documents containing j.

it is correct 78 percent of the time (i.e., a precision of 0.78). This is good considering that, because only 22 percent of the articles in the training sets were labelled as relevant, classifying the articles at random would yield a precision of only 22 percent. Moreover, even with a codebook, deciding whether an article is about environmental policy or not requires a subjective judgement on the part of the annotators. We find that our precision statistics is close to our inner-annotator agreement of 83% (i.e. how much two separate annotators agreed on average when labeling environmental and climate policy articles). If humans cannot perfectly identify relevant articles, we cannot expect our algorithm to do so either. In addition, the algorithm successfully retrieves more than two thirds of the relevant articles (i.e., a recall of 0.67).[15] This is satisfactory given that we are mostly interested in identifying trends in the relative volume of environmental policy news over time, rather than the exact universe of all relevant articles.

Finally, we apply the prediction rule produced by the algorithm on the whole sample. For each of our 459,000 articles, we input its respective tf-idf matrix to the algorithm, which then predicts whether the article belongs to the 'environmental policy' category or not. Our classifier identifies 80,045 relevant articles. Hence, less than 20 percent of all articles from our broad query end up being relevant in our final corpus, which is in line with insights from our manual labeling exercise.

Descriptive statistics

Table 1 displays the features that have the highest weight in predicting whether an article talks about environmental and climate policy or not, according to our classifier. All features are those that one would expect to find in an article about environmental and climate policy. They are a mix of both environmental (i.e. 'energy', 'emissions', 'environmental' or 'climate change') and policy-related terms (i.e. 'obama', 'epa', 'standards', 'federal' or 'regulations'). There are two noteworthy items to keep in mind when interpreting the top-40 discriminating keywords. First, the list of keywords in Table 1 is far from exhaustive and, while many top-scoring words relate to air pollution and climate change as these topics are highly discussed in news articles, we show in Section 3.2 that our classifier also identifies articles on many other topics (e.g. vehicle fuels, water pollution, toxic and hazardous waste). Second, many top-discriminating keywords are not by themselves specific to 'environmental policy' (with the exception of 'epa'). This is

[15] Precision is the fraction of documents identified as relevant by the classifier that were indeed labelled as relevant by the annotators. Recall is the fraction of the relevant documents that are successfully retrieved by the classifier. The F1-score is 0.72 with a precision of 0.78 and a recall of 0.67.

Table 1: Top discriminating keywords for predicting our EnvP index according to the trained SVM classifier.

Word	Weight	Word	Weight	Word	Weight
energy	3.16	crist	1.34	volkswagen	1.09
emission	3.06	air	1.33	refrigerator	1.08
environmental	2.95	ethanol	1.32	utility	1.07
epa	2.24	global warming	1.32	cleanup	1.06
solar	2.17	coal	1.30	federal	1.05
obama	2.05	climate	1.26	car	1.00
clean	1.89	regulation	1.24	penalty	0.99
pollution	1.83	program	1.18	house	0.98
waste	1.67	renewable	1.17	bannon	0.98
warming	1.62	reef	1.15	bill	0.98
recycle	1.47	protection	1.14	mercury	0.97
power	1.45	climate change	1.12	electric	0.96
global	1.38	env. protection	1.10	gasoline	0.94
standard	1.36	clean air	1.10	environment	0.94

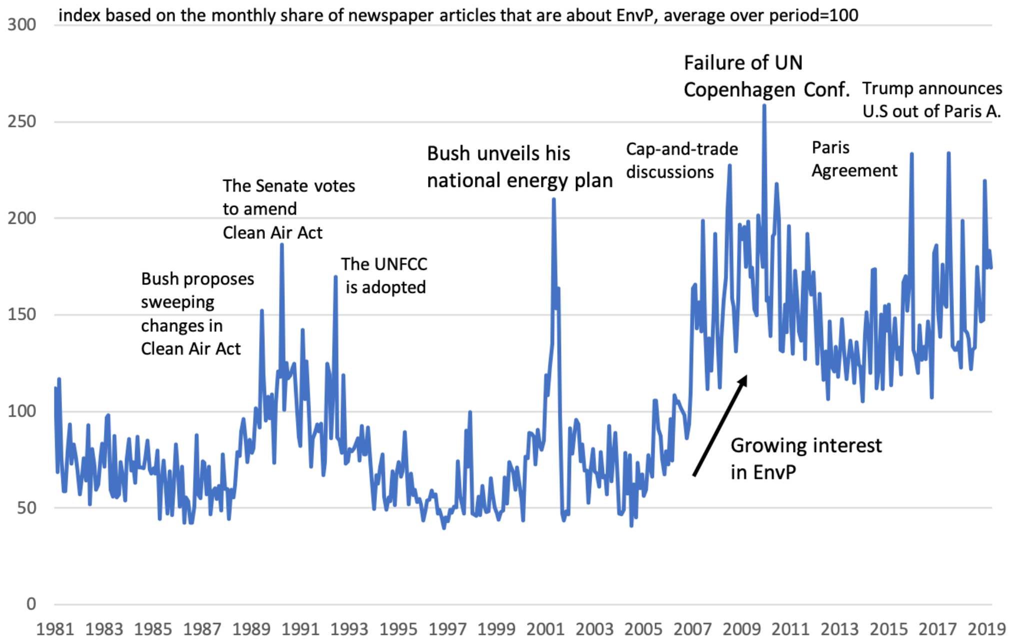
not a major concern given that our initial set of articles includes keywords relating 'climate change and the environment' and 'policy and regulation' within a proximity of 40 characters. Hence, our discriminating keywords help the algorithm to decide which articles are truly about environmental and climate policy within this set.

The SVM algorithm assigns an SVM-score to each article, based on its probability of being classified in the 'environmental and climate policy' category. Table 2 reports excerpts of the five newspaper articles with the highest SVM score. All of these articles are extensively covering environmental policy issues, giving us confidence in our classifier. The first article titled *Environment — Handicapping the Environmental Gold Rush* is a special edition about the green transition and the crucial role that government policies will play in shaping the future of both dirty and clean energy. The second article *In Texas, clean energy set to boom*, describes the ongoing changes in the electricity sector in Texas and the impact of future air pollution regulations.

Table 2: Newspapers articles with the highest SVM-score

Title	Date	Score	Newspaper	Excerpt
Environment — Handicapping the Environmental Gold Rush	Oct 29, 2007	3.55	Wall Street Journal	”The green stampede is on. As a global economy powered by cheap fossil fuel comes under intense pressure to change, corporate executives are racing to stay ahead of the tectonic shift in their world. From Capitol Hill to California and Brussels to Beijing, multinational companies are stepping up their lobbying [...]”
In Texas, clean energy set to boom	Jan 10, 2016	3.54	Dallas Morning News	”While Texas has long been the top state for oil and gas, much more is going on here. In electricity, cleaner-burning natural gas plants are pushing out coal faster than in the rest of the nation, and that’s before the next air pollution regulations kick in.”
Obama Flies to the Nevada Desert to Promote Solar Energy	Aug 25, 2015	3.53	New York Times	”While promoting the benefits of all renewable energy, including wind power, the president focused largely on solar energy, part of an increasingly intense effort to counter global warming by instituting policies to reshape the nation’s energy industry.”
New rule targets pollution from coal	Aug 2, 2015	3.49	Washington Post	”The Obama administration will formally adopt an ambitious regulation for cutting greenhouse-gas pollution on Monday, requiring every state to reduce emissions from coal-burning power plants and putting the country on a course that could change the way millions of Americans get their electricity.”
Environmentalists, Industry Air Differences on Pollution	Oct 17, 1999	3.48	Washington Post	”As a result, environmental groups are pressing states and Congress for specific environmental protections against increased pollution, financial incentives for energy efficiency and renewable energy, and federal pollution guidelines to be part of the overall deregulatory effort.”

Figure 1: EnvP - An index of environmental policy, monthly share, available online https://www.financingcleantech.com/envp-index.



To construct our index, we count the monthly number of articles classified as 'environmental and climate policy' by our SVM algorithm. Since the total amount of news published in newspapers varies over time, we scale the monthly counts of environmental articles predicted by the total monthly volume of news articles in our ten newspapers. Figure 1 plots our environmental policy index (EnvP index). The index is normalized such that its average value over the 1981-2019 period is equal to 100. As shown in Figure 1, our news-based EnvP index is able to capture both trends and spikes in US environmental and climate policy history. We observe more than a three-fold increase in the EnvP index between 2006 and 2009. At the tail-end of this trend, our index identifies two important events which precipitated a fall in both media interest and political will to address climate change, namely 1) the parliamentary debates over the Waxman-Markey bill in April 2009 which failed to introduce a cap-and-trade system and 2) the UN Copenhagen Conference in December 2009 which ended on an unbridgeable North-South divide. Other events, such as the signature of the Paris Climate Agreement in December 2015 and President Trump's announcement of withdrawal from the agreement in June 2017, are labelled in Figure 1.

2.2 Evaluating our news-based environmental policy index

In this section we provide further descriptives and validity checks of our news-based index of US environmental and climate policy.

Comparison with existing indices

We first discuss how our index differs from currently available newspaper-based environmental indices. In Figure 2 we compare our EnvP index with the Climate Change News index created by Engle et al. (2020). The Climate Change News index counts the occurrence of keywords related to climate change in articles published by *The Wall Street Journal* (WSJ). Both indices share similar trends over time, which is reassuring because it means that even though we use different methods, we both broadly capture the same trends.[16] There are, however, notable differences between the two indices. Compared to the Climate Change News index, our EnvP index introduces four novel features. First, we focus specifically on the regulatory and policy framework, i.e. the policy solution to climate change and environmental problems. As a result, policy terms are well represented among the top features of our own classifying exercise, as the words 'obama', 'epa', 'standard', 'regulation' or 'program' in Table 1 show. By contrast, the most important features in the Climate Change News index[17] of Engle et al. (2020) are about the science of climate change and global warming ('carbon', 'emissions', 'temperature', 'atmosphere') and strategies to address it in international agreements ('mitigation', 'adaptation'). Only one policy term – 'protocol' – is present in their top features because of its use in international agreements. Second, we include a much broader set of newspapers across several regions in the US, rather than the more international audience of the WSJ alone. As a result, we are able to capture domestic policy topics both at state and federal level with greater granularity than the Climate Change News index, which reacts a lot to international climate change negotiation events. This is manifest in Figure 2. The Climate Change News index has less pronounced trends and tends to skyrocket mostly during the UN Climate Talks (e.g. adoption of the UNFCCC in 1992, Kyoto protocol in 1997, Copenhagen in 2009, etc). By contrast, while our EnvP index also picks up these events, it is better able to capture other domestic environmental policies that are not directly linked to climate change events, such as

[16] The two series are positively correlated to one another with a correlation coefficient of 0.67 for the monthly series and 0.75 for the quarterly series. If we compare the Climate Change News index to our EnvP index based solely on the WSJ, a fairer comparison, the monthly correlation climbs up to 0.71. See Figure A1 in Appendix A.

[17] See world cloud summary in Figure 1 of Engle et al. (2020)

the Clean Air Act in 1990[18], the failed adoption by the US Congress of a cap-and-trade bill in 2009 or the federal policy support for clean energy companies like Solyndra.[19] Most noticeably, our index better captures the large number of discussions on environmental policy on the US political agenda at both federal and state level between 2008 and 2014, a period where climate change was not necessarily at the forefront (with one notable exception being the failure of the 2009 UN Climate Change Conference in Copenhagen). For instance in January and February 2009, our EnvP index picks up all the discussions about Obama's plan to fight Climate Change and his speech to Congress.[20] Third, we consider a much broader set of environmental concerns than climate change alone. While many environmental policy news indirectly relate to greenhouse gas emissions and climate change (e.g renewable energy, vehicle fuel efficiency, etc), our index also includes articles on other local air pollutants, oil spills, water pollution, toxic and hazardous waste, among others. Finally, a last notable difference is that we provide a more sophisticated methodology than Engle et al. (2020) to identify and classify relevant news with automated text-mining techniques combining both supervised and unsupervised machine learning algorithms.

Given our focus on the regulatory and policy framework, an important question in our analysis is how our index relates to environmental policy stringency. Therefore, as an additional reality check, we compare our EnvP index (12-month moving average) to the OECD's Environmental Policy Stringency Index (EPS) for the United States in Figure 3. The EPS measures the extent to which a country puts an explicit or implicit price on polluting or environmentally harmful behaviour. We see that the indices co-move, with a correlation coefficient of 0.79 between 1990 and 2015. The EnvP index seems nonetheless more sensitive to one-off policy events, such as the energy crisis of 2001.

[18]In April 1990, when the Clean Air Act amendment is passed by the senate our index reached the value of 184, which means it was 84% over its average 1984-2017 level. By contrast, in April 1990, the Climate Change News index was only 7% above its 1984-2017 level.

[19]Solyndra received a $535 million U.S. Department of Energy loan guarantee, and was the first recipient of a loan guarantee under Obama's economic stimulus program, the American Recovery and Reinvestment Act of 2009. The bankruptcy of Solyndra in 2011 has received a lot of attention in the media and was used by Obama's political opponents as an example of wasteful spending under the stimulus program.

[20]We find in our sample that 108 articles mention "Obama" AND "Congress" in January and 88 in February 2009. The typical articles include quotes like : "This week, in his speech to Congress, Mr. Obama made clear that he is ready to spend both to combat climate change and reduce this country's dependence on fossil fuels." or "Attacking climate change through a complex greenhouse gas trading system is a centerpiece of the incoming Obama administration's energy policy". During these months, the EnvP index is at a mean level of 192 (92% above its average). By contrast, these policy discussions are completely missed by the Climate Change News index. It is at 103.4 in January/February, 3.4% above its average.

Figure 2: Comparison with the Climate Change News index from Engle et. al, quarterly share

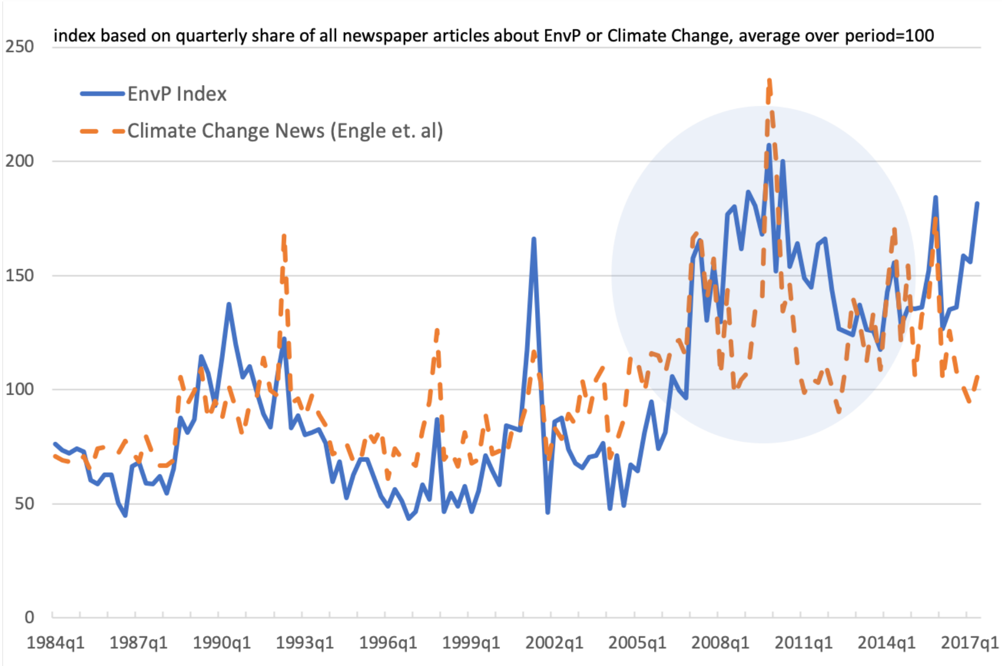
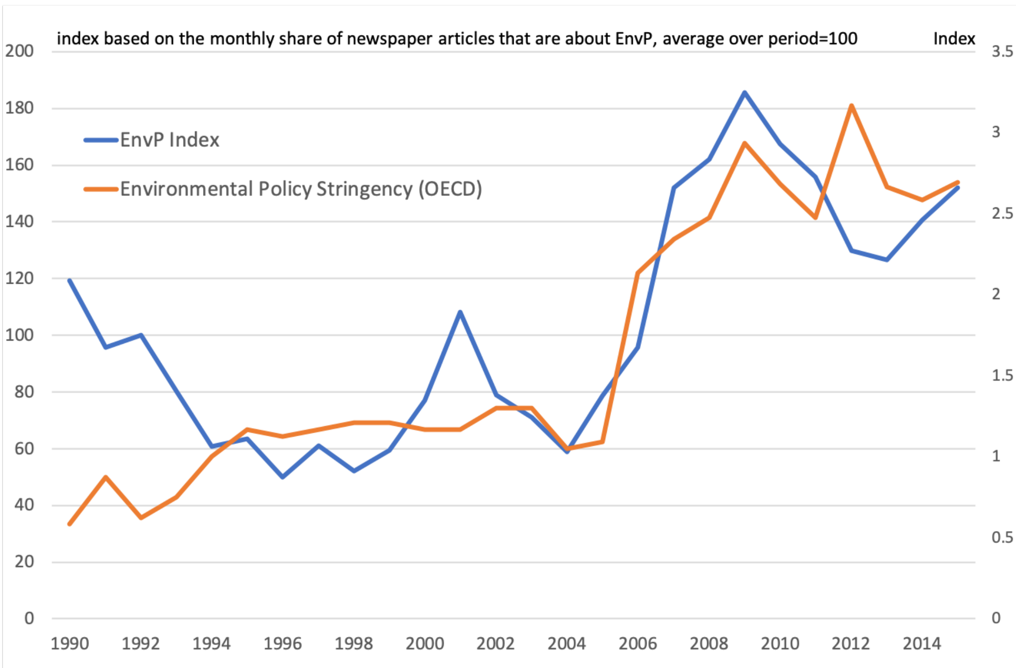


Figure 3: Comparison with EPS, yearly share



Is environmental policy coverage partisan?

A newspaper-based measure of environmental policy should not be overly influenced by the political slant of the newspapers in our sample. To investigate this issue, we divide the newspapers in our sample into two groups based on whether they are more conservative or liberal-leaning.[21]

- Liberal-leaning: *New York Times, Washington Post, San Francisco Chronicle, Tampa Bay Times, San Diego Union Tribune and San Jose Mercury News*
- Conservative-leaning: *Wall Street Journal, Houston Chronicle, Boston Herald and Dallas Morning News*

First, we find that 0.55% of articles in liberal-leaning newspapers are about environmental policy and 0.48% in the more conservative-leaning ones. We plot the EnvP indices produced by the liberal-leaning and conservative-leaning newspapers in Figure 4. The figure shows that the coverage of environmental policy has followed the same trends in these two groups over the past four decades. There are only a few minor exceptions. Notably, liberal-leaning newspapers dedicate more space to environmental policy in the early-months of Trump's presidency than conservative-leaning newspapers. However, as our sample of newspapers is well balanced between liberal and conservative outlets, our general EnvP index averages out the differences. Overall, we observe that political slant does not seem to skew the coverage of environmental policy and is thus not a serious concern for our analysis.

3 Additional Measures of Sentiment and Topic-specific Indices

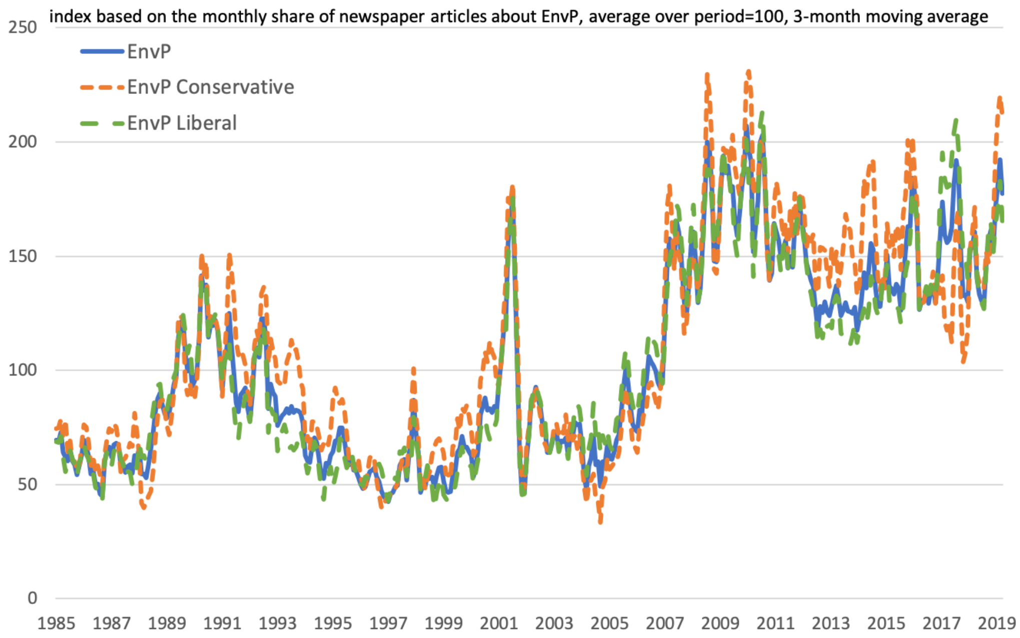
In this section, we introduce two additional types of measures that can be extracted from our main index of environmental and climate policy, namely: 1) a sentiment index and 2) topic-specific indices.

3.1 Sentiment analysis

News about environmental policy may be either positive, negative or neutral. We may be concerned that our EnvP index inaccurately captures negative discussions on environmental and climate policy (due to opposition, protest, rollbacks), giving rise to perceptions of a decline

[21]To determine whether a newspaper is more conservative- or liberal-leaning, we use two external sources: Boston University (`https://library.bu.edu/c.php?g=617120&p=4452935`) and AllSides, a multi-partisan organisation that studies media bias (`https://www.allsides.com/`).

Figure 4: EnvP according to liberal and conservative media



in stringency. Hence, we find it important to control for the sentiment on environmental policy news as conveyed by journalists when we estimate the relationship between our EnvP index and clean investments in Section 4. A sense of optimism (pessimism) in the news could be perceived as increasing (decreasing) policy stringency and growing (declining) opportunities for clean markets. To assess the polarity of our index, we conduct a sentiment analysis following Consoli et al. (2021). The authors develop a novel fine-grained aspect-based methodology that allows for identifying topic specific sentiment around certain keywords, as opposed to merely identifying the general sentiment of a given sentence. The advantage is that, by selecting suitable policy keywords, we make sure to pick up the sentiment actually pertaining to policy, not to any confounding sources of sentiment such as financial markets. A central element in sentiment analysis is the dictionary which assigns each word in the lexicon a sentiment score. In Consoli et al. (2021), the lexicon base is optimized for economic and financial texts.[22]

As it is our aim to measure sentiment pertaining to environmental policy, the algorithm,

[22]We also considered the dictionary-based approach by Loughran and Mcdonald (2011). The downside of this approach for this application is that, as mentioned in Consoli et al. (2021), their dictionary is based on 10-K filings and not on newspaper articles. Moreover, their list of terms for negative sentiment is much longer than that of positive sentiment which inevitably affects the level of sentiment as picked up by the index. However, despite the difference in average sentiment, the trends in the sentiment indices based on Consoli et al. (2021) and Loughran and Mcdonald (2011) are roughly similar with a correlation of 0.46 for the raw index and 0.71 for the six-month moving average. This gives us confidence that the general trends in sentiment are not driven by any methodological particularity. Furthermore, Consoli et al. (2021) show that their methodology outperforms Loughran and Mcdonald (2011) and other common methods in the literature when comparing the predicted labels with a human-annotated sample of texts.

searches around the following terms: 'federal', 'state', 'court', 'treaty', 'summit', 'political', 'administration', 'talks', 'policy', 'congress', 'epa', 'senate', 'regulation', 'rule', 'penalty', 'program', 'house', 'bill', 'protection', 'legislation', 'standard'. The term set was selected from the list of keywords most important for our EnvP classifier as we can be sure that these appear in most of our EnvP articles. For each article, the algorithm identifies sentences where one of the words specified above appears and searches around it to generate a sentiment score per relevant instance. This yields multiple scores per article which all vary between -1 and +1, with zero representing neutral sentiment. Next, we compute the average sentiment score for each article which leaves us with one sentiment score per article.

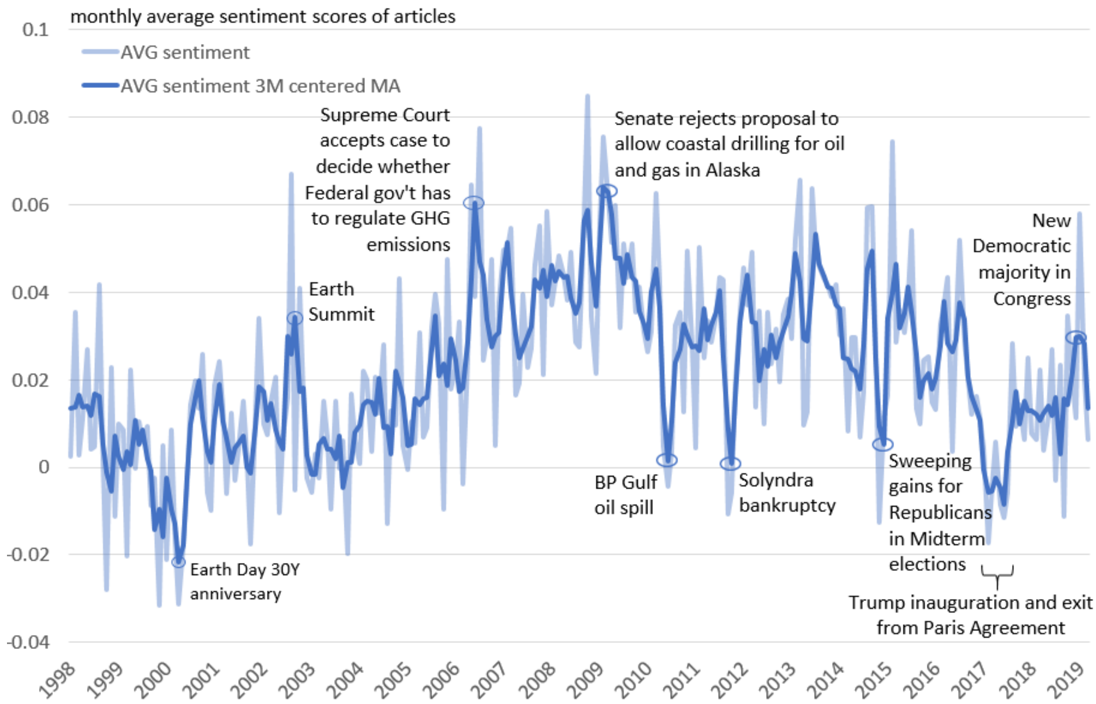
Figure 5 displays our EnvP average sentiment index. Sentiment has been fluctuating around the neutral cutoff of zero until the end of 2002 after which monthly sentiment scores tend to remain positive. The uptick in sentiment in the 2000s is likely driven by a general increase in public discussions on climate change (e.g. through Al Gore's movie 'An Inconvenient Truth', Jacobsen (2011)) and coincided with a strengthening of environmental policy in the US with favourable reporting in the news. Other positive and negative events are labelled in Figure 5. Later dips in sentiment include the 2014 US elections which led to sweeping gains for the Republicans threatening to thwart Obama's climate policy agenda, corresponding to the largest two-month dip in sentiment across the whole sample. Finally, as expected, we also find a major dip in sentiment in June 2017 when Trump announced the US' exit from the Paris agreement.

3.2 Topic-specific indices

Finally, we provide additional descriptive analysis to illustrate that our index captures a vast amount of fine-grained information on various environmental and climate policy topics. We apply topic modeling, an unsupervised machine learning approach, to demonstrate how our index can be decomposed to identify specific environmental policy topics. As an example, Green New Deal policies may include provisions specific to sub-topics on 'automobile emissions' or 'renewable energy'. Unsupervised learning approaches can help discover implicit patterns in the data without researchers imposing any specific structure (such as keywords or a training set). This technique identifies re-occurring word patterns to infer a given number of topics within our corpus of articles.

As a first step, and to limit the number of unique terms included in our analysis, we build a tf-idf matrix of the whole sample of unigrams, bigrams and trigrams included in our 80,045

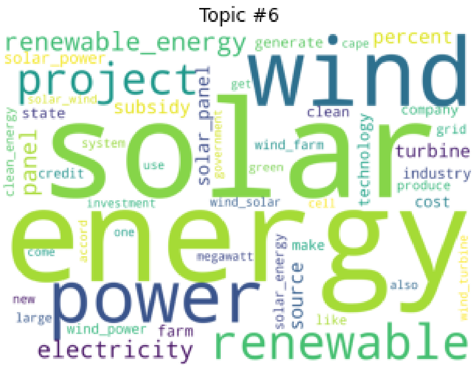
Figure 5: EnvP sentiment index



environmental policy articles and select the 20,000 with the highest tf-idf score.

We then apply Latent Dirichlet Allocation (LDA) topic modeling developed by Blei et al. (2003) and already successfully applied in the economics literature (Hansen & McMahon, 2016; Hansen et al., 2018; Bybee et al., 2020). LDA is a statistical model that views each document as a collection of topics and each topic as a collection of keywords. A given keyword can be attributed to different topics with varying proportions and, likewise, an article can be 80 percent about 'automobile emissions' and 20 percent about 'renewable energy'. We provide detailed information on the LDA algorithm and our methodological choices regarding the number of topics in Appendix C. In our final analysis, we choose to focus on 25 topics pertaining to environmental and climate policy. In order to interpret the topics uncovered by the LDA, we look at the most prevalent words per topic. Figure 6 display the keywords for two exemplary topics using word clouds. The size of a word within a cloud corresponds to its probability of occurring within the topic. The word cloud in Figure 6a is composed of terms such as *energy, solar, wind, power, renewable, electricity, credit, etc.* We label this topic as 'renewable energy'. Similarly, the word cloud in Figure 6b of the combination of words: *united, country, agreement, united_state, world, international, environment, etc.* We label this topic as 'international climate negotiations'.

Figure 6: Word clouds



(a) Renewable energy policy

(b) International climate negotiations

In a similar fashion, we interpret additional topics as for instance air pollution, water pollution, automobile industry, waste and recycling, green buildings, cleanups and courts, etc.[23] Figure 7 illustrates the evolution of topics over time. In the 80s and 90s, the most important news topics were 'cleanup and courts', 'water pollution' and 'air pollution'. More recently, 'renewable energy' and 'climate change' have gained increasing attention.

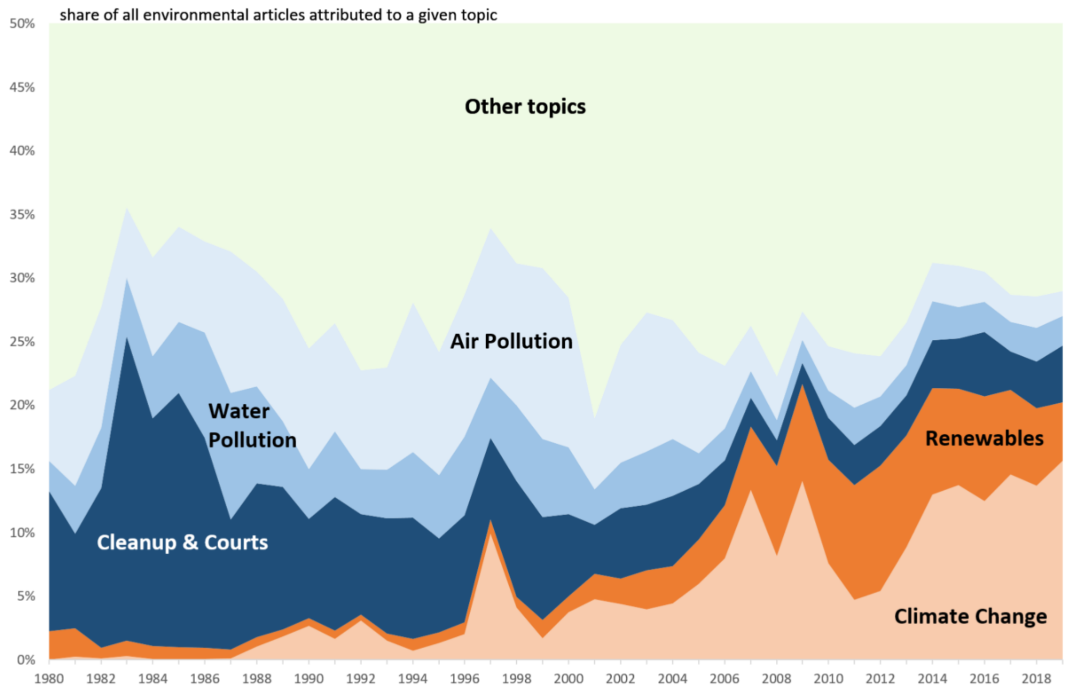
To construct topic-specific indices, we count the number of articles attributed to a given topic and scale it by the total volume of newspaper articles in our sample. As before, we scale the indices so that their average corresponds to 100 over the 1981-2019 period. Figures 8 and 9 plot the evolution of the index for the sub-topic of renewable energy policy and international climate negotiation, respectively.

Figure 8 shows that policy discussions on renewable energy rise significantly after 2005, which corresponds to the period of implementation and strengthening of renewable portfolio standards in many US states, particularly during the Obama presidency. The announcements of the Green New Deal in February 2009 and the Clean Power Plan in August 2015 aiming to reduce carbon dioxide emissions from electrical power generation by 32 percent by 2030 relative to 2005 levels marked key milestones toward this agenda. Other specific issues reported in news articles are indicated directly on Figure 8.

Figure 9 shows that our international climate negotiations index correctly captures discussions around major international climate summits. It identifies events such as the Rio Earth Summit in June 1992, the Kyoto Protocol in December 1997 and the Bonn Climate Change Conference in July 2001 which was the first international meeting after President Bush had exited the Kyoto Protocol in March 2001. In more recent years, the Copenhagen Climate

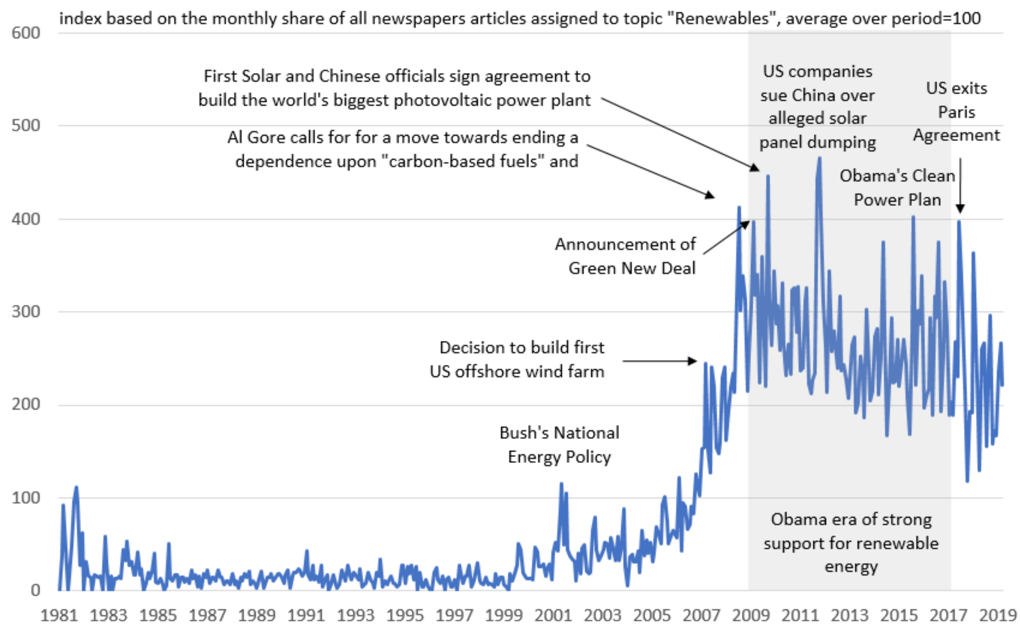
[23]The full list of topics is provided in Table C1 in Appendix C.

Figure 7: Evolution of news on selected environmental and climate policy topics over time



The figure shows how relative news on environmental policy topics vary over time. In the 80s and 90s, the most important topics were 'Cleanup and Courts', 'Water Pollution' and 'Air Pollution'. More recently, 'Renewable Energy' and 'Climate Change' have become central topics in the media.

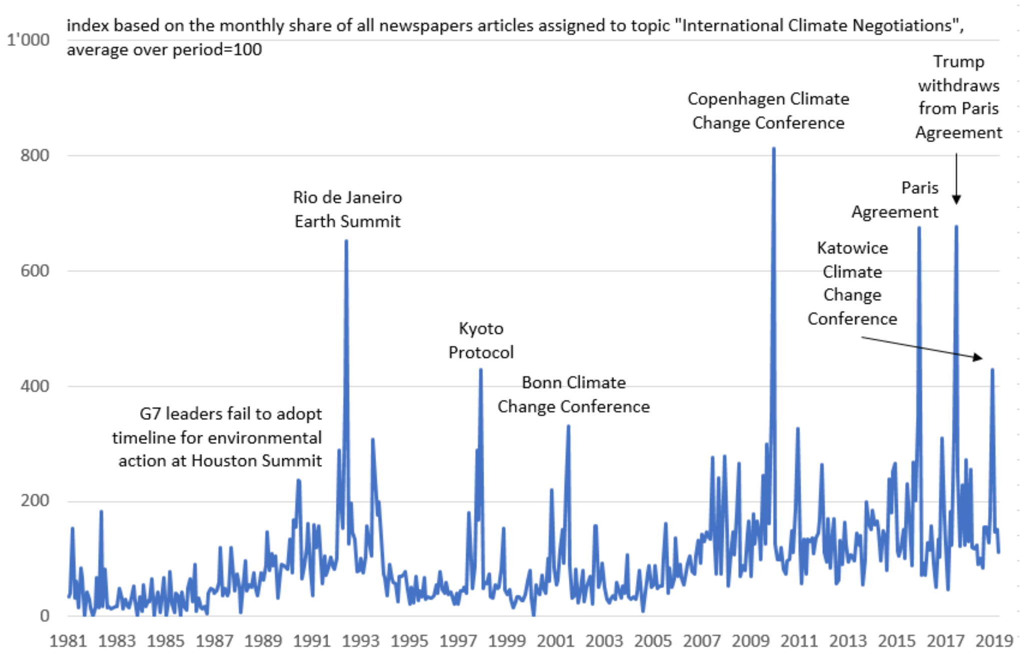
Figure 8: Index - Renewable energy policy



Change Conference in December 2009, the Paris Agreement in December 2015 and the COP24 in Katowice are all picked up as salient events by our index.

Finally, beyond this selective set of illustrations, we believe that there are many additional

Figure 9: Index - International climate negotiations



powerful applications of our EnvP index. Our index and topic model can for instance easily be combined with keywords to identify a more fine-grained set of policies. We provide in Figures A2 and A3 in Appendix A further illustrations of a news-based index for the state of California and of an index depicting the environmental policy coverage around ExxonMobil.

The purpose of our topic modeling exercise is mainly to illustrate how our index can be disaggregated into sub-topics that can be of use to researchers in future empirical work. We believe that the advantage of unsupervised machine learning is that the choice of topics is not pre-determined by researchers. Given the large number of texts and large geographical coverage of our set of newspaper articles, our topic model provides much richer insights on environmental policy topics than what we could have ourselves identified via a more structured approach. Yet, given the limited scope of the present analysis, an in-depth discussion and analysis of the various additional topics is left for future work. Instead, in the remainder of the analysis, we will only consider in our robustness analysis how our sub-index on renewable energy policy (EnvP-RE) relates to investments in renewable energy, since specific data are available on this topic.

4 Environmental Policy and Investments in Clean Technologies

We now turn to the central part of our analysis, documenting the meaningful association between our EnvP index and clean markets. Conceptually, we expect that a rise in the volume of news on environmental and climate policy signals growing policy support and thus increasing opportunities for clean products, technologies and firms. In other words, we expect a rise in our EnvP index to be associated with an increase in investments in clean markets. Conversely, an increase in our index could signal vanishing opportunities for dirty markets, so we may also expect our EnvP index to be negatively associated with investments in dirty and polluting products and firms. In our empirical analysis, we consider two proxies for investments, namely venture capital finance and stock markets valuation, and conduct our analysis both at the micro level using firm-level estimations and at the aggregate level using VAR models. As establishing causality is challenging,[24] our firm-level regressions aim to validate the meaningful association between our EnvP index and financial investments in startups and firms most exposed to environmental policy, as defined by their sector of activity or emissions levels. Our analysis using VAR models provides insights on the dynamic relationship between our news-based EnvP index and investments in clean energy markets at the sectoral level and might potentially capture additional channels (e.g. entry and exit).

4.1 Environmental Policy and Firm-Level Clean Investment Decisions

4.1.1 VC investments across industries

We first examine how our EnvP index is associated with the probability that a startup will receive VC funding. So far, the empirical literature on the determinants of VC funding for cleantech startups is quite limited (Criscuolo & Menon, 2015; Popp et al., 2020; van den Heuvel & Popp, 2022). Yet, one of the main advantage of VC data for our purpose is that information on deals are available at a high frequency.

We obtain data on VC funding rounds between January 1998 and March 2019 for US startups from the Crunchbase database and aggregate these funding rounds into a firm-quarter panel dataset.[25] We also extract firms' industry and founding date as well as all the information

[24]First, environmental regulations news may respond to expectations of future clean market growth and technological advancements. Investors might therefore anticipate policy news which could lead to a downward bias in our estimate. Second, both our index and clean investments may be affected by additional omitted variables. For instance, growing environmental awareness leading to a shift of consumers' preferences towards clean goods would likely raise both environmental policy news and investments in clean technologies leading to an upward bias in our estimate.

[25]We focus on series A to J financing, involving firms founded after 1985. This represents around 75,000 different funding rounds. Due to the panel nature of our dataset, we observe startups over long period of times and therefore should avoid including inactive startups. These inactive startups can either have gone bankrupt or not be in need of early-stage financing anymore and as such have a probability equal to zero to receive VC funding. We therefore classify any firm that fails to secure new financing within the three-year time span after its last round of financing as inactive after this three-year mark.

related to the funding rounds (i.e., date, amount, series) from Crunchbase. Finally, we include GDP data from the U.S. Bureau of Economic Analysis, the Federal Reserve effective funds rate and the West Texas Intermediate crude oil spot prices. Excluding firm-quarter observations containing any missing information–including those where the firm is classified as inactive–we obtain 1,056,221 firm-quarter observations on 35,637 unique startup firms.

We differentiate startups by their exposure to environmental and climate policy: more precisely, we expect to find a positive association between our EnvP index and the probability of VC funding for startups classified in cleantech industries, while we expect no significant relationship for other startups. Cleantech startups belong to Crunchbase's 'Sustainability' industry group and represent 4% of overall VC deals, while clean energy startups in clean energy, battery, renewable energy, wind energy, energy storage and solar industries represent only 2.4% of all VC deals.We estimate whether startups that are classified as cleantech or clean energy are significantly more responsive to our EnvP index than startups in other sectors using ordinary least squares (OLS) as follows:[26]

$$\begin{aligned} VC_{i,t+1} = & \alpha + \beta_1 EnvP_t + \beta_2 EnvP_t \cdot Cleantech_i + \beta_3 Controls_{i,t} + \\ & \beta_4 TimeTrend_t + \gamma_{quarter/year/industry/state/series} + \epsilon_{i,t} \end{aligned} \quad (1)$$

where i indexes the firm, t the quarter. We use two different measures of VC investments as our dependent variable: a funding dummy, *Funded*, and the logarithm of the total amount of funding a startup receives during a quarter, *Amount*, conditional on $Funded = 1$. β_1 and β_2 are the coefficients of our two main variables of interest. β_1 identifies the association between a rise in our EnvP index on non-cleantech startups, while β_2 on the other hand captures the relationship with cleantech (clean energy) startups that we expect to be most exposed to environmental and climate policy.

We control for the following variables that could be confounding our results. First, our EnvP index is likely affected by business cycles effects, as environmental concerns might take a backseat role during a crisis. We therefore control for economic activity and capital availability by including the annual growth of U.S. GDP and the Federal Reserve effective funds rate. Second, we include the log of the oil price as it is both an important actor in the environmental policy debate and actual investment decisions. In some specifications we include the output

[26]Our results are robust to using Probit.

of our sentiment analysis on our EnvP index. This allows us to control for the positive (or negative) content of the news.

We also include a set of variables and fixed effects to absorb variation that is unrelated to environmental and climate policy but may nonetheless affect our results, including, firm i's age in quarter t – set as missing before founding date and if it is inactive – as well as a time trend, and in some specifications an industry time trend. We also use firm, quarter, year and series funding round fixed effects.[27] The firm fixed effect control for firm-level unobservables such as firm's performance. The quarter fixed effects are used to account for seasonality in the data.[28] The other fixed effects also allow us to control for unobserved variables common to all startups in a given year or funding round. Finally, we cluster standard errors at the startup firm level to correct for potential serial correlations in the error term.

We first focus on the relationship between our EnvP index and VC investment in cleantech using Equation (1). We present the regression results in Table 3, first using the probability of getting funded in the next quarter (Q + 1) as the dependent variable. Using column (1) we can see that while a rise in our EnvP index is associated with a marginally lower chance of receiving funding for the average startup, it has a strong positive relationship with cleantech startups' probability of receiving funding in the next quarter.[29] To illustrate the size of the effect, a doubling of environmental policy media coverage from one quarter to the next is associated with an increase in the probability of receiving funding of 1.4 percentage points.[30] While this might seem like a small increase, the average probability that a cleantech startup will be funded next quarter in our sample is only 6.2 percent. Therefore a doubling of environmental policy media coverage is actually associated with a 23 percent increase in a cleantech startup's probability of receiving funding next quarter. Column (2) shows that news with a positively-toned sentiment are associated with more VC deals in cleantech. After correcting for sentiment, the EnvP index continues to be significant, indicating that both the volume and sentiment of articles matter to investors. In column (3), we see that our result holds when our EnvP index is orthogonalized to the Climate Change News index by Engle et al. (2020). Keeping the EnvP

[27] The series funding rounds dummies capture whether the investment is a series A, series B all the way up to Series J.

[28] Additional estimations including quarter-year fixed effects provide similar results.

[29] Furthermore, we find in additional results presented in Appendix that this effect persists in the next quarters but declines over time. Table D2 in Appendix D shows the results for the probability of getting funded in Q+2 and Q+3. The coefficient of the interaction term between our EnvP index and cleantech startups gradually declines over the quarters.

[30] We obtain this number by doing the following calculation $(-0.00538 + 0.0253) \cdot 0.693$, given that a doubling in a logged variable implies an increase by 0.693.

index constant, the Climate Change News index has no significant relationship with cleantech VC investments. This suggests that what matters for VC investments in cleantech are news about climate change policies, rather than about climate change itself. Columns (4) and (5) use the natural logarithm of the amount received in dollars, given that they received funding next quarter, as the dependent variable. We find that a one percent increase in our EnvP index is associated with a 0.6 percent increase in the amount received by cleantech startups. This remains robust when controlling for the sentiment index.

The relationship between the EnvP index and VC investments in non-cleantech startups is either insignificant (columns (4) and (5)) or negative (column (1)). We do not have a priori expectations on how our EnvP index could affect VC investments in non-cleantech startups. While we could expect non-cleantech startups to remain unaffected, we could also expect a negative association between our EnvP index and VC financing if environmental regulations are wide-ranging and costly for the average firm.[31] In any case, the baseline results highlight the fact that the EnvP index is disproportionately associated with cleantech firms.

As an additional robustness check and to illustrate the application of topic-specific subindices, we consider how our sub-topic index on renewable energy policies (EnvP-RE) relates to the financing of startups active in renewable and fossil fuels industries. We expect renewable energy startups to be more affected by EnvP-RE news than other startups. By contrast, we expect VC funding of fossil fuel startups to have either no or a negative relationship with the EnvP-RE index. Table 4 displays our results. Results on the interaction terms show that a rise in our EnvP-RE index is associated with both a higher probability for renewable energy startups to secure funding and a higher amount per funding. At the same time, the EnvP-RE index has no significant relationship with VC investments in fossil fuels startups.

4.1.2 Firm-level stock returns

Next, we examine how our EnvP index relates to firm-level stock returns in panel estimations, drawing on the emerging literature in environmental economics looking at how environmental policy signals are reflected in firms stock valuations (Kruse, Mohnen, & Sato, 2020; Mukanjari & Sterner, 2018; Barnett, 2019).[32]

[31] There is a large body of empirical work looking at the relationship between environmental regulations and the profitability of manufacturing firms (Cohen & Tubb, 2018), which confirms the considerable heterogeneity, with studies find either an insignificant, positive or negative impact of environmental regulations on firms' profits and productivity.

[32] In finance, a recent literature explores whether high-polluting firms exposed to carbon risks are receiving a risk compensation in the form of higher stock returns. Bolton and Kacperczyk (2020) find for instance that

Table 3: Baseline results - EnvP index and VC investments in cleantech

	(1) Funded (Q+1)	(2) Funded (Q+1)	(3) Funded (Q+1)	(4) Amount (Q+1)	(5) Amount (Q+1)
Log EnvP index	-0.00538***	-0.0110***	-0.00652***	-0.0168	-0.0188
	(0.00194)	(0.00194)	(0.00217)	(0.0484)	(0.0491)
Log EnvP x Cleantech	0.0253***	0.0212***	0.0228***	0.467***	0.379***
	(0.00479)	(0.00485)	(0.00547)	(0.147)	(0.142)
Log Sentiment index		-0.00901***			-0.0220
		(0.000659)			(0.0136)
Log Sentiment x Cleantech		0.00936***			0.278***
		(0.00216)			(0.0628)
Log Climate Change News Index			-0.00759***		
			(0.00204)		
Log Climate Change News x Cleantech			0.00514		
			(0.00645)		
Firm FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Industry-time trend	Yes	Yes	Yes	Yes	Yes
Series FE	Yes	Yes	Yes	Yes	Yes
Observations	1056221	1056221	935517	57319	57319
Firms	35637	35637	34218	28297	28297
R^2	0.006	0.006	0.007	0.118	0.119

The table presents results of an OLS regression. The sample period is January 1998 and March 2019. The dependent variable in Columns (1) to (3) is a dummy variable that indicates whether firm i received VC funding next quarter. In Columns (4) and (5), the dependent variable is the logarithm of the amount received, conditional on having received funding. Controls include age, oil price, a time trend, GDP and the fed fund rate. Standard errors are clustered at the company level. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We start our analysis by collecting monthly total return indices for a sample of around 1400 firms across various industries listed on the US stock exchange from January 2004 to March 2019 from Datastream. We also extract the monthly safe interest rate from the website of Kenneth French[33] and compute monthly continuously compounded log returns at the firm level as $r_{i,t} = \ln\left(\frac{p_{i,t}}{p_{i,t-1}}\right)$. We use the excess returns above the safe rate r_f, i.e. $r^e_{i,t} = r_{i,t} - r_f$, as our dependent variable. In the estimation, market returns are controlled for on the right-hand side by five market risk factors (Fama & French, 2015).

When working with stock price data, we may be particularly worried that investors may anticipate and react very quickly to movements in our EnvP index. To mitigate this, we follow

firm-level carbon emissions significantly and positively affect the firm's stock returns, suggesting that forward-looking investors are seeking compensation for holding the stocks of high-polluting firms. Most of the literature on risk-compensation in finance is however concerned with cross-sectional analysis, i.e. trying to distinguish how various firms' characteristics (such as carbon emissions) affect the cross-section of firms. By contrast, our focus is different since we examine how a rise in our EnvP index is associated to *within* firm variation in stock returns over time in panel data settings.

[33] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 4: Robustness - EnvP-RE index and VC investments in clean energy

	(1) Funded (Q+1)	(2) Amount (Q+1)
Log EnvP-RE index	0.00657***	-0.00378
	(0.00119)	(0.0293)
Log EnvP-RE index x Renewables startup	0.0134***	0.616***
	(0.00312)	(0.106)
Log EnvP-RE index x Fossil fuels startup	-0.00497	0.00434
	(0.00488)	(0.121)
Firm FE	Yes	Yes
Quarter FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes
Industry-time trend	Yes	Yes
Series FE	Yes	Yes
Observations	1056221	57319
Firms	35637	28297
R^2	0.006	0.119

The table presents results of an OLS regression. The sample period is January 1998 and March 2019. The dependent variable in Column (1) is a dummy variable that indicates whether firm i received VC funding in the next quarter. In Column (2), the dependent variable is the logarithm of the amount received, conditional on having received funding. Other controls include age, oil price, a time trend, GDP and the fed fund rate. Standard errors are clustered at the company level. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Brogaard and Detzel (2015) in considering 'innovations' in our EnvP index (i.e. its unanticipated component) by extracting the residuals from an AR(7) model of our monthly series of EnvP as follows:

$$\epsilon_t^{EnvP} = EnvP_t - \left(\hat{\phi}_0 + \sum_{k=1}^{7} \hat{\phi}_t * EnvP_{t-k} \right) \tag{2}$$

Standard tests confirm that this series is white noise and has no autocorrelation.[34] We standardize this measure to have a mean of zero and a unit standard deviation.[35]

Just as before, we differentiate firms according to their exposure to environmental policy. To proxy for such exposure, we use firm-level scope 1 CO_2 emissions[36] collected from publicly disclosed data sources at annual frequency from S&P Trucost Limited. As environmental reg-

[34]Breusch–Godfrey test for higher-order serial correlation, Durbin's alternative test for serial correlation and the Portmanteau (Q) test for white noise.

[35]In the same fashion, we extract residuals from an AR(6) model from our monthly EnvP sentiment index.

[36]Scope 1 emissions are direct emissions from production, as opposed to scope 2 emissions which are indirect emissions from consumption of purchased electricity, heat or steam.

ulations gain prominence in the news, we expect investors to divest from high-emission firms, leading to lower stock returns. Since CO_2 emissions and thus firm exposure to environmental policy may endogenously respond to current or anticipated environmental policy, we use the (standardized) fixed mean of CO_2 emissions over the sample period in our baseline estimation. We consider the following panel estimation:

$$
\begin{aligned}
r^e_{i,j,t=m} =& \alpha + \beta_1 \epsilon^{EnvP}_{t=m} + \beta_2 \text{CO2 Emissions}_{i,t=y} + \beta_3 \text{CO2 Emissions}_{i,t=y} * \epsilon^{EnvP}_{t=m} + & (3)\\
& \beta_4 \text{Risk Factors}_{t=m} + \beta_5 \text{Firm controls}_{i,t=y} + \beta_6 \text{Time Trend}_{j,t=y} + \gamma_{\text{i}} + & (4)\\
& \varepsilon_{i,t=m} & (5)
\end{aligned}
$$

where i, j, t indicate firm, industry and time (with m denoting month and y denoting year), respectively. CO2 Emissions$_{i,t}$ proxies firm-level environmental policy exposure by the fixed mean of CO_2 emissions over the period. $\epsilon^{EnvP}_{t=m}$ represents the monthly EnvP innovations. Risk Factors$_t$ is a vector containing the monthly market risk factors MKTRF, SMB, HML, RMW and CMA from the 5-factor Fama-French asset pricing model (Fama & French, 2015) obtained from the website of Kenneth French.[37] In addition, $X_{i,t}$ is a vector of firm-specific characteristics, namely (i) firm size as log(market cap), (ii) a measure of firm profitability as log(return on assets), (iii) a measure of firm leverage as log(total debt/total equity) as well as (iv) log(dividends per share). Table E1 in Appendix E provides summary statistics of all variables used in the analysis. Finally, we include an industry-year time trend in all our specification to control for time-varying factors specific to industries, such as technological progress, as well as firm fixed effects to control for structural and time-invariant differences in stock returns at the firm level. We cluster standard errors at the firm level to control for serial correlation of the error terms.

Table 5 presents the results of our baseline estimation. Column (1) includes firm fixed effects, Fama-French risk factors and industry-year trends. We add firm controls in columns (2)-(7), which reduces the sample to about 600 firms. Columns (3)-(5) test different policy exposure measures such as quartiles of (fixed mean) CO_2 emissions, CO_2 emission intensities and pre-sample CO_2 emissions. In column (6) we add controls for innovations in our sentiment index. In column (7) we add the Climate Change News index by Engle et al. (2020) as a control.

[37] We deflate all financial variables by annual GDP collected from the database of the St. Louis Fed.

Table 5: Baseline results - EnvP index and Excess Stock Returns

	(1) Exc. ret.	(2) Exc. ret.	(3) Exc. ret.	(4) Exc. ret.	(5) Exc. ret.	(6) Exc. ret.	(7) Exc. ret.
EnvP index	0.0022***	0.0027***	0.0051***	0.0027***	0.0014***	0.0019***	0.0004
	(0.0003)	(0.0003)	(0.0011)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
EnvP index × AVG CO_2 Emissions	-0.0003*	-0.0004***				-0.0004***	-0.0004***
	(0.0001)	(0.0001)				(0.0001)	(0.0001)
Quartile of CO_2 emissions=2 × EnvP			-0.0016				
			(0.0012)				
Quartile of CO_2 emissions=3 × EnvP			-0.0030***				
			(0.0012)				
Quartile of CO_2 emissions=4 × EnvP			-0.0030***				
			(0.0011)				
EnvP index × CO_2 Emission Intensity				-0.0003**			
				(0.0001)			
EnvP index × Pre-sample CO_2 Emissions					-0.0006***		
					(0.0002)		
EnvP Net Sentiment index						-0.0043***	
						(0.0005)	
EnvP Net Sentiment index × AVG CO_2 Emissions						0.0001	
						(0.0002)	
Climate Change News index							0.0004
							(0.0004)
Climate Change News index × AVG CO_2 Emissions							0.0001
							(0.0001)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Risk factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	91,531	49,143	49,143	49,143	34,079	49,143	38,762
Firms	1,400	614	614	614	262	614	557
R^2	0.54	0.72	0.72	0.72	0.77	0.72	0.77

The table presents results of an OLS regression. Standard errors are clustered at the firm level. The dependent variable corresponds to excess returns as continuously compounded monthly returns in excess of the safe rate. Emission measures refer to scope 1 fixed average (AVG) CO_2 emissions or emission intensity at the firm level. Firm controls include size as log(market capitalization), profitability as log(return on assets), leverage as log(total debt over total equity) and log(dividends per share). Risk factors include the market risk factor, SMB, HML, RMW and CMA. We consider 'innovations' in the EnvP index, EnvP Net Sentiment index and Climate Change News index as the residuals from an AR(7), AR(6) and AR(4) process, respectively. These are standardized to a mean of zero and unit standard deviation. The CO_2 emission measures are standardized in the same way. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Across all specifications, we find that our coefficient of interest, i.e. the interaction term between our EnvP index and CO_2 emissions, has the expected sign and is highly significant at the 1 percent level. There is a negative association between our EnvP index and the stock returns of high-emission firms with greater exposure to environmental policy. Quantitatively, firms with CO_2 emissions one standard deviation above the sample mean experience an associated differential drop in excess returns of around 4 basis points concurrently to a one-SD EnvP innovation, as shown in column (2). For a firm with average exposure[38], however, news on environmental regulations tend to be positively associated with stock returns, as indicated by the positive coefficient on EnvP.[39] In column (3), we find that there are highly significant differences between the least polluting quartile and the two most polluting quartiles, with firms in the two highest polluting quartile experiencing the largest relative drop in excess returns

[38]In the estimation, we standardize the CO_2 emissions variable, with mean zero and unit SD.
[39]We investigate this further by controlling for news sentiment in column (6).

when EnvP rises relative to the least polluting quartile of firms.

The association between EnvP and stock returns is robust to a variety of adjustments to our baseline specifications. Using emission intensity or pre-sample emissions as a policy exposure measure does not materially alter the size or significance or our coefficient of interest, as shown in columns (4)-(5). Moreover, our results are robust to controlling for net sentiment about environmental policy news, as shown in column (6). These findings underline that our EnvP index has is negatively associated with the stock returns of firms bearing a greater exposure to environmental policy regardless of the sentiment of EnvP news. For a firm with average exposure, however, an increase in positively toned news about environmental policy is associated with a drop in excess returns, as indicated by the negative coefficient on the EnvP net sentiment variable. As far as positive sentiment reflects a strengthening of environmental policy, this is aligned with the intuition that more stringent environmental regulations may be costly for most (average) firms. However, we do not find evidence of a significant difference for firms most exposed to environmental policy. Finally, column (7) shows that our results are robust to including the Climate Change News index by Engle et al. (2020). This finding further confirms that our EnvP index relates significantly to financial markets even after controlling for broad climate change news, in line with the notion that investors are sensitive to policy signals.

4.2 Environmental Policy and Aggregate Clean Investments

Having looked at firm-level estimations, we now consider the association between our index and aggregate investments in the clean energy sector at the macro level. By contrast to within-firms decisions, this may capture additional channels (e.g. entry and exit) of the dynamic relationship between investments and our news-based index of US environmental policy.

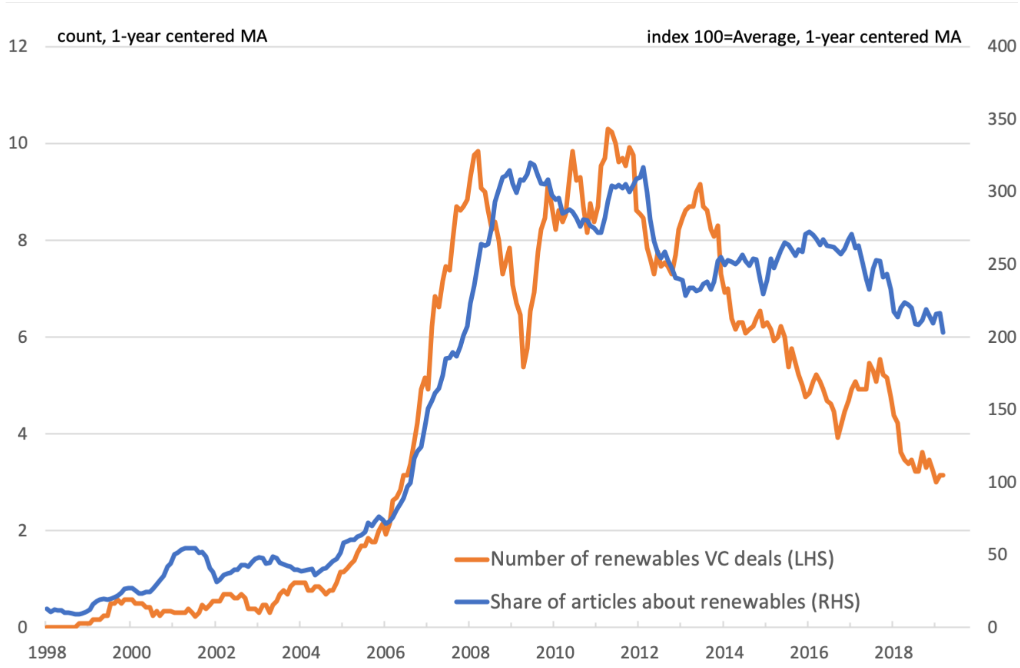
4.2.1 Aggregate cleantech venture capital deals

We extract data on the monthly number of venture capital deals in renewable energy (which includes solar, wind, hydro and geothermal) over the January 1998 - March 2019 period from the i3 Cleantech Group database.[40] Since we focus on renewable energy, we use our index on 'renewable energy policy', as this is likely the most relevant for investors. Figure 10 plots our EnvP-RE index together with the aggregate monthly number of VC deals in renewable energy.

[40] This database provides information on early-stage financing of 11,620 US cleantech startups (seed, series A, series B and growth equity) tracked over time by the Cleantech Group.

Both series share a similar trajectory since the beginning of the 2000s, only diverging during the global financial crisis in 2008-2009 and over the 2015-2017 period.

Figure 10: Evolution of number of renewable energy VC deals and EnvP-RE news index, monthly



Our baseline VAR specification includes the following controls, all at monthly frequency: 1) oil prices as the West Texas Intermediate crude oil spot price from the St-Louis FED, 2) market risk captured by the Federal Reserve effective funds rate from the Board of Governors of the Federal Reserve System, 3) aggregate economic activity using Markit's U.S. monthly real GDP index[41] and 4) a linear time trend. We include three lags of all variables, based on lag selection criteria. Table D1 in Appendix D provides summary statistics of the variables in our sample.

We conduct standard unit roots tests and use the monthly first difference of the following series, the log of oil prices, the log of GDP and the Federal funds rate, because these are not stationary in levels. As we can reject the presence of a unit root for the number of VC deals and the EnvP-RE news index using the Phillips–Perron test, we keep these two variables in levels in our preferred specification.[42] In order to recover orthogonal shocks we use the following Cholesky ordering: EnvP-RE news index, $\Delta \ln(\text{oil price})$, $\Delta \ln(GDP)$, Δ effective Fed funds rate, VC deals in renewable energy.

[41]For our robustness analysis below using Californian data, we use the Federal Reserve Bank of Philadelphia's coincident economic indicator, which includes non-farm payroll employment, the unemployment rate, average hours worked in manufacturing and wages and salaries.

[42]We expect VC deals which take several months to close to be more strongly correlated with the level of environmental policy in the media rather than its monthly evolution.

Figure 11: Estimated effect of a shock in EnvP-RE news on the number of renewable energy venture capital deals

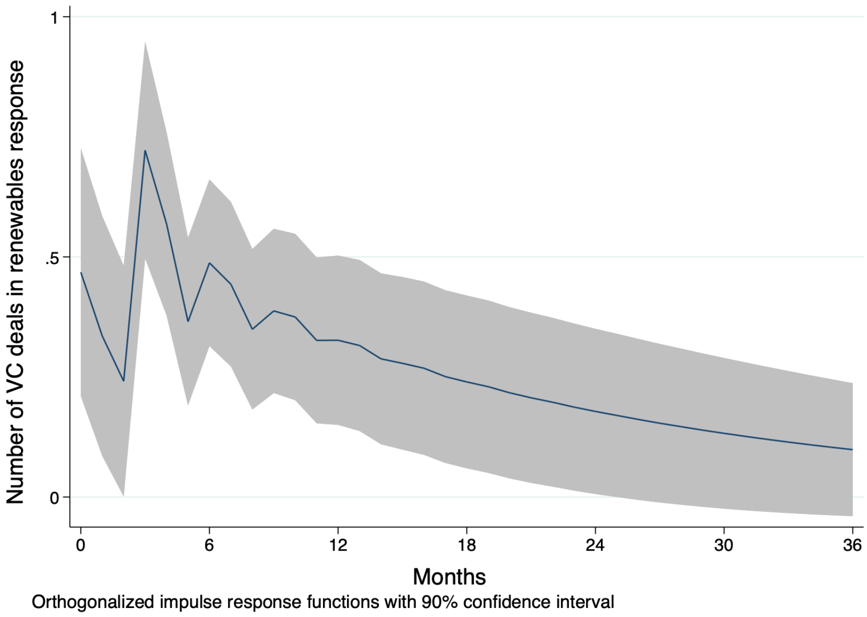


Figure 11 displays the model-implied impulse response function of the number of VC deals in renewable energy to a shock in our news-based EnvP-RE index. We see that a one standard deviation increase in our index is associated with about 0.6 more VC deals in the medium term. While this effect is moderate in size, it still represents a nearly 15% increase in the average monthly number of VC deals in renewable energy (i.e. 4.2 between January 1998 and March 2019). Interestingly, Figure 11 confirms the results from the firm-level analysis; it takes several months for changes in the EnvP index to be reflected in clean energy VC deals. We show that this positive relationship between the EnvP-RE news index and VC investments in renewable energy is robust to varying specifications on Figure D1 in Appendix D.

4.2.2 Aggregate clean energy stocks

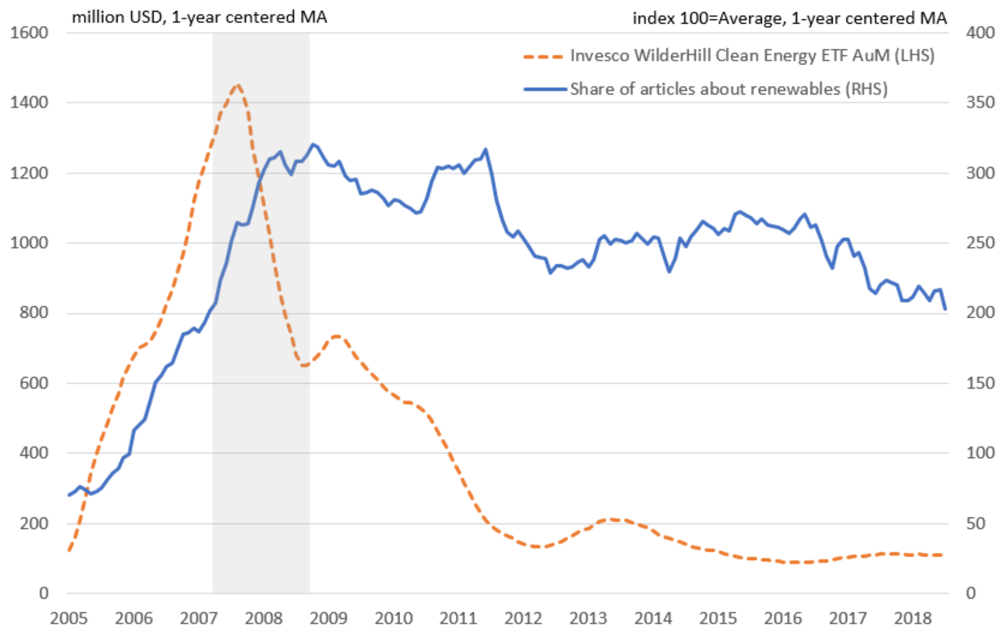
Next, we investigate the dynamic relationship between our news-based index and aggregate clean energy stocks. Specifically, we examine how the assets under management (AuM) of the Invesco WilderHill Clean Energy Exchange Traded Fund (PBW-ETF), tracking the portfolio of 52 US renewable energy companies, is associated with our index.[43] Considered as the main

[43]The reason we are focusing on an ETF are twofold. First, one cannot directly invest in a market index. Second, it allows us to analyze the investment behaviour of less sophisticated investors who are more likely to learn something new from journal articles because retail investors are the main participants in the ETF market. A caveat of using the assets under management of an ETF is that they may be driven by fund flows or by changes in the value of the underlying assets. Therefore, it is a measure of demand for renewable energy stocks likely

benchmark clean-energy index, the PBW-ETF is widely used in the energy economics literature (Kyritsis & Serletis, 2019; Sadorsky, 2012; Kumar et al., 2012). We extract this series from Datastream. Again, given the focus on renewable energy, we use our specific EnvP-RE index to measure news on renewable energy policy.

Figure 12 plots the monthly sub-index on renewable energy policy together with the assets under management of the PBW-ETF for the period of March 2005 to March 2019. The figure shows that the co-movement patterns of the series vary substantially over time, with a high co-movement before the Global Financial Crisis (GFC) but less so during and after it.[44]

Figure 12: Evolution of PBW-ETF and EnvP-RE news index, monthly. The shaded area corresponds to the recession following the Global Financial Crisis from December 2007 to June 2009.



Our baseline VAR specification includes the monthly assets under management of the PBW-ETF, EnvP-RE news and other controls as in (Kyritsis & Serletis, 2019; Sadorsky, 2012; Kumar et al., 2012), namely: 1) oil prices, as the US West Texas Intermediate crude oil spot price, 2) technology stocks, using the NYSE Arca Technology Index (PSE), and 3) market risk captured by the Federal Reserve effective funds rate. We exclude the recession associated with the GFC

suffering from some measurement error.

[44]The correlation of the annual centered moving average of PBW-ETF AuM and our news index between 2005 and 2007 is very high at 0.9. During the recession caused by the Global Financial Crisis (GFC), officially dated by the NBER from December 2007 - June 2009, PBW-ETF AuM take a dip, while policy news about renewable energy remain at elevated levels amid the announcement of the Green New Deal and Obama's era of strong support for renewable energy. The correlation of PBW-ETF AuM and our news index during this time period is at -0.9. The post-GFC period is marked by a much lower co-movement of the PBW-ETF AuM with our EnvP-RE policy news index at 0.7 (0.6).

(December 2007 - June 2009) from the analysis.

As before, we run a series of unit root tests (Augmented Dickey–Fuller, Phillips–Perron and Kwiatkowski–Phillips–Schmidt–Shin tests). Accordingly, we use the PSE and oil prices in monthly log differences, PBW-ETF AuM and the effective Fed funds rate in monthly differences and, although tests are less conclusive in this case, we use the EnvP-RE index in monthly log differences in our baseline – so we consider changes in the growth rate of our EnvP index. We include one lag for all variables as suggested by standard tests and recover orthogonal shocks by imposing the following Cholesky ordering: $\Delta \ln$(EnvP-RE index), $\Delta \ln$(oil price), ΔFederal Reserve effective funds rate, $\Delta \ln$(PSE), Δ(ETF-PBW).[45] Summary statistics are provided in Table E2 in Appendix E.

Figure 13: Estimated effect of a shock to the growth rate of the EnvP-RE news index on the change in assets under management of the PBW exchange traded fund

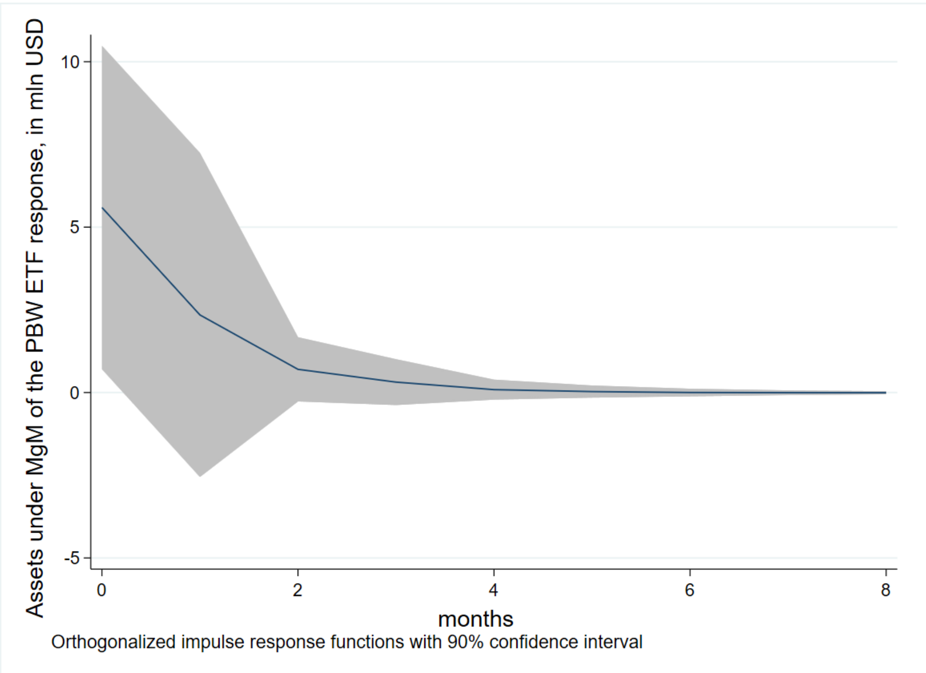


Figure 13 shows that a one-standard deviation shock to the growth rate of our EnvP-RE index is associated with an additional increase of 5 million USD in assets under management of the PBW-ETF. While this effect seems rather small, it still represents a 125% increase in the average monthly change in AuM of the PBW ETF (i.e. about 4 mln USD between April 2005 and March 2019). The result is broadly in line with the previous literature which finds a quantitatively small dynamic relationship between investor sentiment in renewable energy,

[45] Akaike information criterion (AIC), Final Prediction Error (FPE) and Hannan-Quin information criterion (HQIC)

as measured by the Google Trends Search Volume Indices and Tweets, and clean energy stock returns (Reboredo & Ugolini, 2018; Song et al., 2019).[46] Figure E1 in Appendix E shows that our results are robust to various other specifications.

5 Conclusions

Quantifying fine-grained information on environmental and climate policy over several decades has often proven difficult. We apply text-mining techniques to newspapers archives to develop the EnvP index, a novel news-based index of US environmental and climate policy over the 1981-2019 period. The index captures the evolution of the relative share of news articles discussing environmental and climate regulations over the last four decades. We perform several reality checks showing that our index accurately captures trends and peaks in the historical evolution of US environmental and climate policy and co-moves with the stringency of the regulatory framework in a meaningful way. We further look at how our EnvP index relates to financial investments in clean markets. Our results provide a range of empirical evidence corroborating that our news-based measure of environmental and climate policy has a meaningful association with clean investments as proxied by venture capital financing and stock returns – both in firm-level panel estimations and VAR models. More specifically, a doubling of environmental policy news is associated with a 26% increase in the likelihood than an average cleantech startup receives funding. Conversely, a 1 SD increase in (an innovation in) our EnvP index is associated with a loss of about 4 basis points in excess returns for the most polluting firms. Furthermore, we find in VAR models that a shock in our sub-index on renewable energy policy is associated with an increase in the number of clean energy deals at the macro level and an increase in the assets under management of the main clean energy exchange-traded fund. Our analysis showcases how newspaper archives combined with machine learning algorithms for text classification can be exploited to retrieve a vast and diverse amount of information on environmental and climate policy. In addition, the algorithms provide a much improved methodology, compared to simpler information retrieval using keywords. We illustrate how the index can be further exploited to

[46] Investors in renewable energy markets may instead be more responsive to factors that move technology stocks than to environmental regulation. Sadorsky (2012), for instance, point out that renewable energy companies tend to behave similarly to high-tech companies because their success hinges on very specific technologies. Consistent with this hypothesis, we find that the PSE and PBW-ETF AuM have a positive association of about double the size of the link between EnvP-RE and PBW-ETF. Moreover, the link between oil prices and PBW-ETF AuM is about the same size as the one between EnvP-RE and PBW-ETF, in line with the notion that rising oil prices trigger a substitution towards renewable energy technologies (Kumar et al., 2012; Sadorsky, 2012).

build many additional indicators, providing information on sentiment (i.e. the tone of articles) and on sub-topics such as ‘renewable energy policy’ and ‘international climate negotiations’ among others. Despite news being a noisy signal for the underlying state of environmental and climate regulations, our findings across many specifications persistently show that our EnvP index has a meaningful association with clean markets, in a manner that is consistent with environmental policy helping to promote clean markets.

We see several potential applications and extensions for future research. First, an immediate direction for future work is to examine how our EnvP index can assist the financial community by providing an improved quantification of transition (policy) risks in the context of climate change and the low-carbon transition. Second, we see a lot of opportunities for future research and policy analysis in filtering out and exploiting further the wealth of information about environmental regulations contained in the EnvP index, such as the unfolding of the policy process (announcements, delays, revisions), specific policy features (implementation, target groups, compensation schemes, enforcement) or the political context (opposition, controversies, actors), which are typically difficult to quantify and track over time. As an illustration, we build on the EnvP index in follow-up work to classify the subset of news articles pertaining to ‘environmental policy uncertainty’ (Noailly, Nowzohour, & van den Heuvel, 2022). More broadly, we hope that our index can help researchers to progress towards quantifying causal impacts of specific features of environmental regulations, for instance by combining our index with event studies or quasi-natural experiments.

Another worthwhile area of research using our EnvP index would be to examine the importance of media coverage and policy communication for the effectiveness of environmental and climate policy. This could be done for instance by comparing similar policies with different (exogeneously driven) media coverage. Such analysis could provide useful insights on how policymakers can coordinate investor beliefs by communicating about their environmental and climate policy agenda in a clear and credible manner, akin to central banks coordinating inflation expectations through forward guidance.

Finally, there are many ways in which our methodology could be extended to develop additional indexes of state versus federal environmental and climate regulations, natural resource policies (e.g. forest, fishery) or using improved text classification methods. We hope that researchers will consider many of these avenues in future work.

References

Albrizio, S., Kozluk, T., & Zipperer, V. (2017). Environmental policies and productivity growth: Evidence across industries and firms. *Journal of Environmental Economics and Management*, *81*, 209–226.

Baker, S., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Barnett, M. (2019). *A run on oil: Climate policy, stranded assets, and asset prices* (PhD Thesis). The University of Chicago.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, *90*(1-2), 1–36.

Baylis, P. (2020). Temperature and temperament: Evidence from Twitter. *Journal of Public Economics*, *184*, 104–161.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *Journal of Machine Learning Research*, *3*, 993–1022.

Bohr, J. (2020). Reporting on climate change: A computational analysis of U.S. newspapers and sources of bias, 1997–2017. *Global Environmental Change*, *61*, 102–038.

Bolton, P., & Kacperczyk, M. (2020). *Do Investors Care about Carbon Risk?* (No. w26968).

Botta, E., & Kozluk, T. (2014). Measuring environmental policy stringency in OECD countries: A composite index approach. *OECD Economic Department Working Papers*(1177).

Boykoff, M. T., & Boykoff, J. M. (2007). Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum*, *38*(6), 1190–1204.

Brogaard, J., & Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, *61*(1), 3–18.

Brunel, C., & Levinson, A. (2016). Measuring the stringency of environmental regulations. *Review of Environmental Economics and Policy*, *10*(1), 47–67.

Brüggemann, M., & Engesser, S. (2017). Beyond false balance: How interpretive journalism shapes media coverage of climate change. *Global Environmental Change*, *42*, 58–67.

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2020). *The structure of economic news* [NBER Working Paper No 26648].

Chen, X., Li, L., Xu, G., Yang, Z., & Kitsuregawa, M. (2012). Recommending related microblogs: A comparison between topic and wordnet based approaches. In *Twenty-Sixth AAAI Conference on Artificial Intelligence.*

Cohen, M. A., & Tubb, A. (2018, April). The Impact of Environmental Regulation on Firm and Country Competitiveness: A Meta-analysis of the Porter Hypothesis. *Journal of the Association of Environmental and Resource Economists*, *5*(2), 371–399.

Cole, M. A., & Fredriksson, P. G. (2009). Institutionalized pollution havens. *Ecological Economics*, *68*(4), 1239–1256.

Consoli, S., Barbaglia, L., & Manzan, S. (2021). *Fine-grained, aspect-based sentiment analysis on economic and financial lexicon* [SSRN Working paper].

Criscuolo, C., & Menon, C. (2015). Environmental policies and risk finance in the green sector: Cross-country evidence. *Energy Policy*, *83*, 38–56.

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, *9*(1), 24.

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189.

Dugoua, E., Dumas, M., & Noailly, J. (2022). Text-as-data in environmental economics and policy. *Review of Environmental Economics and Policy*, *16*.

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.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, *114*(2), 280–316.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, *78*(1), 35–71.

Ghanem, D., & Smith, A. (2020). What are the benefits of high-frequency data for fixed effects panel models? *Journal of the Association of Environmental and Resource Economists*, 710–968.

Greenstone, M. (2002). The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. *Journal of political economy*, *110*(6), 1175–1219. (Publisher: The University of Chicago Press)

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*, S114-

S133.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, *133*(2), 801–870.

Jacobsen, G. D. (2011). The Al Gore effect: An Inconvenient Truth and voluntary carbon offsets. *Journal of Environmental Economics and Management*, *61*(1), 67–78.

Johnstone, N., Haščič, I., Poirier, J., Hemar, M., & Michel, C. (2012, June). Environmental policy stringency and technological innovation: evidence from survey data and patent counts. *Applied Economics*, *44*(17), 2157–2170.

Kahn, M. E., & Kotchen, M. J. (2011). Business cycle effects on concern about climate change: The chilling effect of recession. *Climate Change Economics*, *2*(03), 257–273.

Keller, T. R., Hase, V., Thaker, J., Mahl, D., & Schäfer, M. S. (2020). News media coverage of climate change in India 1997–2016: Using automated content analysis to assess themes and topics. *Environmental Communication*, *14*(2).

Kruse, T., Mohnen, M., Pope, P., & Sato, M. (2020). *Green revenues, profitability and market valuation: Evidence from a global firm level dataset* [Unpublished manuscript]. Grantham Research Institute on Climate Change and the Environment, London School of Economics.

Kruse, T., Mohnen, M., & Sato, M. (2020). *Are financial markets aligned with climate action? New evidence from the Paris Agreement* [Working Paper No 333]. Grantham Research Institute on Climate Change and te Environment, London.

Kumar, S., Managi, S., & Matsuda, A. (2012). Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis . *Energy Economics*, *34*(1), 215–226.

Kyritsis, E., & Serletis, A. (2019). Oil Prices and the Renewable Energy Sector. *The Energy Journal*, *40*(01), 337–364.

Kölbel, J. F., Leippold, M., Rillaerts, J., & Wang, Q. (2020). *Ask BERT: How regulatory disclosure of transition and physical climate risks affects the CDS term structure.* (Swiss Finance Institute Research Paper No. 21-19)

Loughran, T., & Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Manela, A., & Moreira, A. (2017, January). News implied volatility and disaster concerns. *Journal of Financial Economics*, *123*(1), 137–162.

Muehlenbachs, L., Sinha, E. N., & Sinha, N. R. (2011). Strategic release of news at the EPA. *RFF Discussion Paper*.

Mukanjari, S., & Sterner, T. (2018). *Do markets trump politics? Evidence from fossil market reactions to the Paris Agreement and the U.S. election* [Working Paper in Economics No. 728]. University of Gothenburg.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *The American Economic Review*, *95*(4), 46.

Noailly, J., Nowzohour, L., & van den Heuvel, M. (2022). *Does Environmental Policy Uncertainty Hinder Investments Towards a Low-Carbon Economy?* [CIES Research Paper 76].

Popp, D., Pless, J., Haščič, I., & Johnstone, N. (2020). *Innovation and entrepreneurship in the energy sector.* (NBER Working Paper No 27145)

Reboredo, J. C., & Ugolini, A. (2018). The impact of Twitter sentiment on renewable energy stocks. *Energy Economics*, *76*, 153–169.

Sadorsky, P. (2012). Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, *34*(1), 248–255.

Sautner, Z., van Lent, L., Vilkov, G., & Zhang, R. (2020). Firm-level climate change exposure. *SSRN Electronic Journal*.

Shapiro, J. M. (2016). Special interests and the media: Theory and an application to climate change. *Journal of Public Economics*, *144*, 91–108.

Song, Y., Ji, Q., Du, Y.-J., & Geng, J.-B. (2019). The dynamic dependence of fossil energy, investor sentiment and renewable energy stock markets. *Energy Economics*, *84*, 104–564.

van den Heuvel, M., & Popp, D. (2022). *The Role of Venture Capital and Governments in Clean Energy: Lessons from the First Cleantech Bubble* (NBER Working Paper No. 29919).

A Additional tables and figures: EnvP index

Table A1: Newspaper list

Newspaper	Available since	% in Dow Jones	% in Query
New York Times	1 June 1980	22.5%	21.9%
Washington Post	6 January 1982	15.3%	17.9%
Wall Street Journal	13 June 1979	9.8%	12.8%
Houston Chronicle	2 February 1985	13.8%	11.1%
Dallas Morning News	18 January 1984	10.8%	7.5%
San Francisco Chronicle	4 January 1985	6.2%	7.4%
Boston Herald	26 July 1991	5.0%	2.7%
Tampa Bay Times	11 June 1986	11.5%	12.3%
San Jose Mercury News	2 January 1994	3.4%	3.9%
San Diego Union Tribune	31 December 2010	1.7%	2.4%

Figure A1: EnvP - WSJ versus Engle et. al

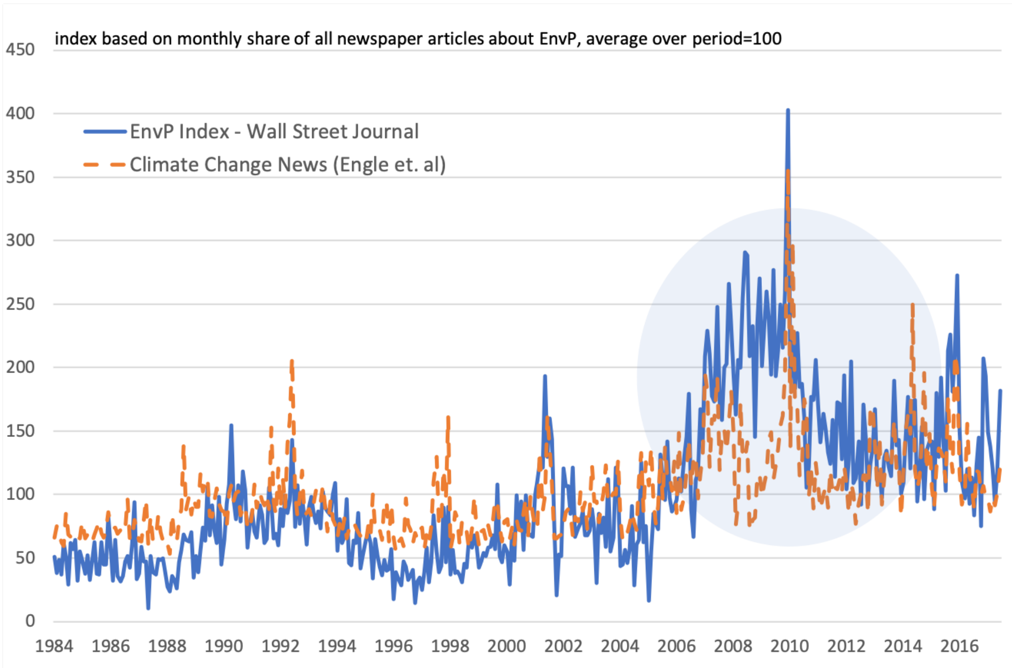
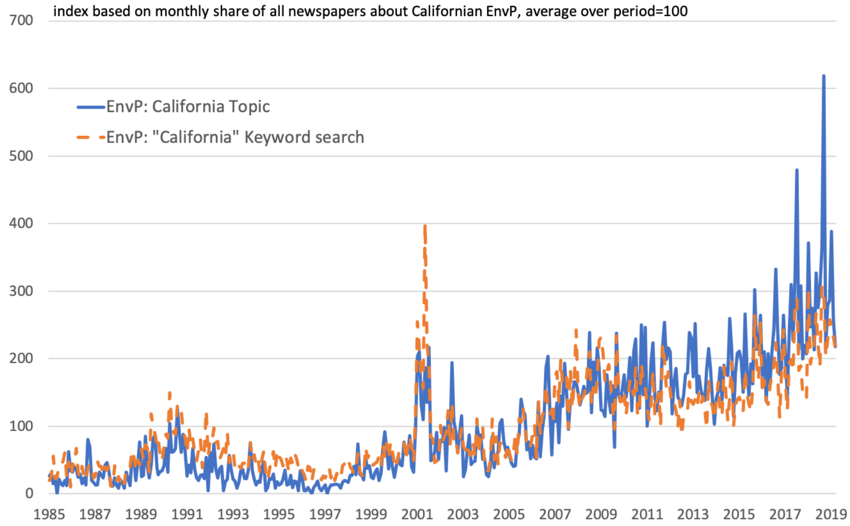
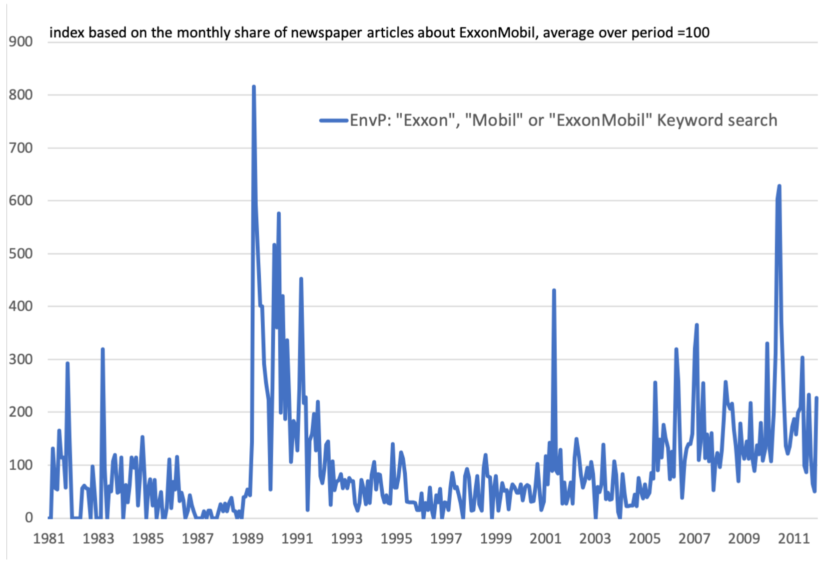


Figure A2: California EnvP indices



The figure plots two alternative versions of a sub-EnvP index for the state of California: 1) a version in which we select articles from the 80,045 articles composing our EnvP index using the "California" keyword, 2) a version in which we plot the index resulting from our California topic in the topic modeling exercise. Both approaches capture similar trends.

Figure A3: ExxonMobil index



The figure plots the evolution of a sub-EnvP index capturing discussions around ExxonMobil in environmental policy news, using the combination of keywords 'Exxon', 'Mobil' or 'ExxonMobil' into our set of EnvP articles. This index correctly identifies the main events around 1) the Exxon Valdez oil spill in March 1989, which led to the implementation of the Oil Pollution Act in the following year, 2) the US withdrawal from the Kyoto protocol and debates around a letter from Exxon to Bush to outcast IPCC's chairman in February/March 2001, 3) Exxon's attacks on the Clean Air Act in 2006, 4) the request from Exxon's shareholders to disclose potential economic impacts of the Paris accord on the company in May 2017.

B The Support Vector Machine (SVM) Algorithm

Given a training set $T = ((x_j, y_i), j = 1..n)$ where x_j are the input variables (i.e. the features and their tf-idf score in each article) and y_i is the corresponding output value (i.e. the assigned label for each article), SVM methods will fit a model to this training set, for a given set of parameters. The algorithm builds on the idea that the numerical representation of each text/article are data points in a multivariate space of features. Based on the word content of each article the algorithm aims to find two hyperplanes separating the two classes of data (i.e., environmental policy articles and the rest). The classifier maximizes the distance between these hyperplanes, also called the margin. Articles or vectors that lie on one of these hyperplanes, i.e. particularly ambiguous articles that were hard to classify, are called support vectors. The decision boundary, that separates relevant from irrelevant articles is the hyperplane that lies at the mid-point of the margin.

Choosing the optimal hyperparameters

We choose the linear kernel function for its best performance. On a more technical note, we rely on a GridSearch function to set up the hyperparameters adapted to our classification model. This procedure is simply an exhaustive search through a subset of the hyperparameters available for the model (the kernel, the regularization parameter, the penalty parameter, gamma, and the class weight). Using this function we can find the optimal combination of hyperparameter values for our model.

Evaluating the performance of the classsifier

In order to evaluate the performance of our classifier, we estimate its out-of-sample performance via tenfold cross-validation. After randomly segmenting the training sets into ten sub-samples, the tenfold cross-validation approach consists in estimating the model on nine of the sub-samples and testing its out-of-sample properties on the tenth one. The procedure is then repeated for every possible permutations of the samples. We obtain a quantification of the performance of the algorithm, which is an average over repeated estimations of five ten-fold cross-validations using different random seeds.

C Topic modeling

Latent Dirichlet Allocation (LDA) is a generative statistical model of a corpus made of D documents — newspaper articles — and V unique terms. This topic model estimates K topics each of which is a distribution $\beta_k \in \triangle^V$ over all the unique terms V present in our articles.

Text pre-processing

We use the pre-processed corpus of 80,045 articles identified in Section 2 and do some additional noise filtering. We tested the robustness of our preprocessing steps with the preText package in R (Denny & Spirling, 2018). The package computes a preText score for a range of text pre-processing specifications indicating whether any given specification is likely to yield 'unusual' results with respect to alternative ways to preprocess. We find that none of our initial preprocessing steps is significantly at risk of leading to unusual results. We include mono-, bi- and trigrams and use a tf-idf approach to filter out words that are either too rare or too common.

Selecting the optimal number of topics

The choice of the number of topics in the model is a critical step. Choosing too low a number tends to result in broad topics that miss the less prevalent but nonetheless important subtopics whereas choosing too high a number can lead to topics that are excessively narrow.

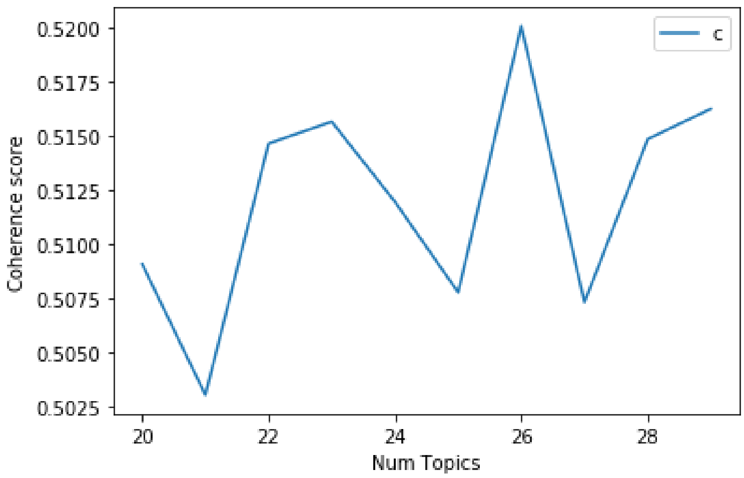
First, we compute the coherence score, a measure indicating how semantically similar the top features of any given topic are to one another. Figure C1 reports coherence scores for LDA models with K ranging from 20 to 30. We favor this intermediate number of topics because while a higher number of topics tends to improve the statistical goodness-of-fit, it also makes the topics less easy to interpret. Next, we inspect the different outputs of these models to determine which one has the clearest and most interesting topics. We settle on a model with $K = 26$, the maximum in Figure C1.

This model precisely identifies multiple topics we are particularly interested in such as policy discussions related to renewable energy, automobile emissions regulations and international climate negotiations. However, out of the 26 topics, one has no interpretable meaning and as a result we discard it as noise and do not take it into consideration when we assign topics to our articles. Table C1 shows our 25 topics ranked from most to least prevalent in our sample.

Table C1: Topic interpretation and classification (ranked by size).

Topic	#	Topic	#	Topic	#
Climate Change	19	Oil & Gas production	15	Vehicle Fuels	12
EPA & Federal Gov.	5	Intl. Climate Negotiations	18	Waste & Recycling	26
Cleanups & Courts	17	Texas	11	Green Buildings	25
Government Budgets	3	Renewables	6	North-East Region	8
Air Pollution	9	Env. Conservation	4	Offshore Oil Drilling	7
Congress & Policy	13	Water Pollution	1	Nuclear Power	21
Businesses & Investments	22	Climate Science	16	Coal Industry	10
Presidents & Campaigns	23	California	14		
Power & Utilities	24	Automobile Industry	2		

Figure C1: Topic coherence



To attribute topics to articles we use the fact that LDA models each article as a distribution over different topics so that each article d has an attribution to topic k, θ_d^k, in percentage. We could simply count all articles with $\theta_d^k > 0$ for any given topic. However, we aim to get rid of 'noisy articles', those whose attribution to a topic is below a critical threshold, $\theta_d^k < \theta_d^{k\min}$. Alternatively, we could simply choose to only pick one dominant topic per article. However, newspaper articles arguably talk about more than one relevant sub-topic of environmental policy and we aim to capture details of the U.S. policy debate. In addition, we believe that contemporary observers may draw information on policy issues, even if it is only mentioned as a side topic. To strike a middle ground, we use a cutoff α of 10%, meaning that only articles associated with any given topic with more than 10% probability, $\theta_d^k > 0.1 \ \forall \ d, k$ are counted.

D Additional statistics and results: venture capital

Descriptive statistics - firm-level estimations and VAR

Table D1 reports descriptive summary statistics for the main continuous variables used in our study of VC investments for the period between January 1998 and March 2019. The first panel shows the variables that are common to both the VAR and Panel analysis, taken from our monthly VAR dataset. The second panel displays statistics about the monthly number of VC deals in the United States that we use in our VAR analysis. Finally, the last panel reports the VC-related variables that we use in our panel analysis. The discrepancy between the number of observations reported in the last panel and in Table 3 comes from the fact that, in our analysis, we drop all the observations where the age variable is negative as it most likely indicates an error in the data. In other words we drop around 30,000 observations that were recorded before the startups' official founding date.

Table D1: Summary statistics for venture capital analysis

	Obs.	Mean	Std. Dev	Min	Max
Environmental Policy Indices:					
EnvP Index	255	119.75	47.34	40.89	258.55
EnvP-RE Index	255	166.28	119.35	0.00	465.94
Economic Control Variables:					
YoY GDP Growth	243	2.19	1.75	-4.92	5.64
Oil price (WTI)	255	57.68	28.59	11.35	133.88
Fed Funds Rate	255	2.09	2.11	0.07	6.54
VAR: VC Variables					
Number of clean energy VC deals	255	15.58	11.25	0.00	48.00
Number of renewables VC deals	255	4.17	4.07	0.00	19.00
Panel: VC Variables					
Number of VC deals, per firm-quarter	1089760	0.06	0.24	0.00	3.00
VC amount raised (in mio), if funded	65061	12.16	36.71	0.00	3500.00
Age when funded (in years)	63989	5.18	4.40	0.00	33.50

Firm-level estimation: robustness analysis

Table D2: Relationship between the EnvP index and VC investment in cleantech over the quarters

	(1) Funded (Q+1)	(2) Funded(Q+2)	(3) Funded (Q+3)
Log EnvP index	-0.00538***	-0.000357	0.00620***
	(0.00194)	(0.00195)	(0.00198)
Log EnvP x Cleantech startup	0.0253***	0.0217***	0.0138***
	(0.00479)	(0.00459)	(0.00485)
Firm FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Industry-time trend	Yes	Yes	Yes
Series FE	Yes	Yes	Yes
Observations	1056221	1036050	1015286
Firms	35637	35637	35637
R^2	0.006	0.005	0.004

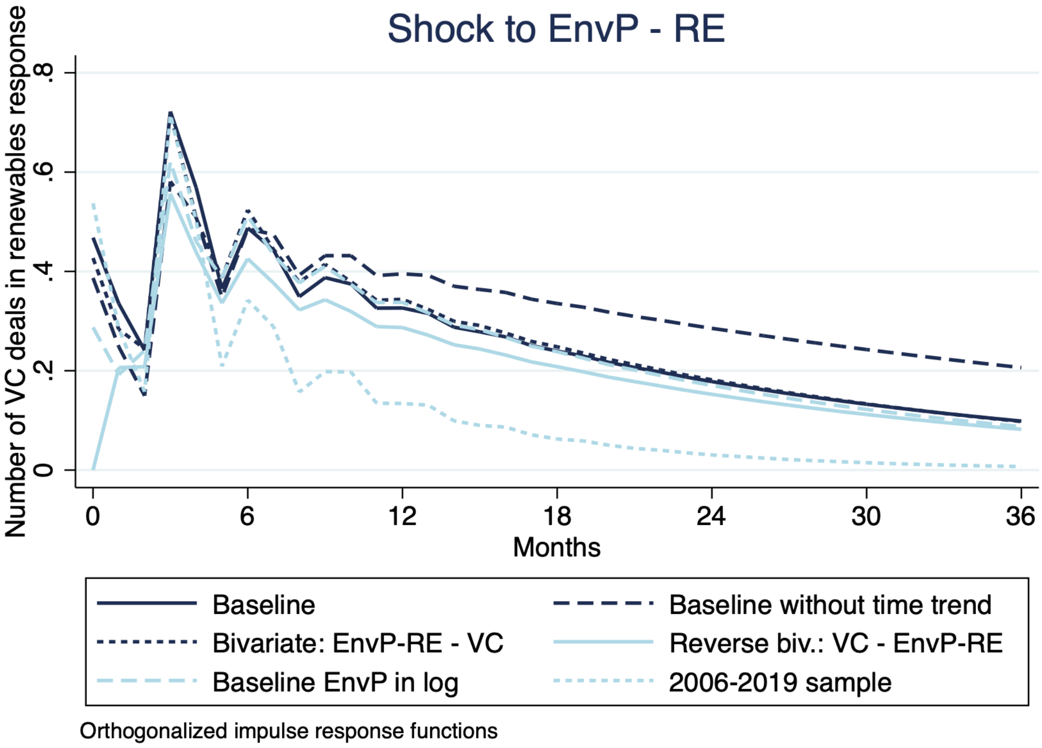
The table presents results of an OLS regression. The sample period is January 1998 and March 2019. The dependent variable in Columns (1)-(3) is a dummy variable that indicates whether firm i received VC funding in Q+1 to Q+3. Controls include age, oil price, a time trend, GDP and the fed fund rate. Standard errors are clustered at the company level. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

VAR: robustness analysis

We provide further robustness analysis in Figure D1. Our baseline response is very similar to the response from a bi-variate VAR model with EnvP-RE and our VC data, as well as the response from this bivariate model with reversed ordering. The responses from VAR models fitted on variables in levels and on a shorter sample from 2006 to 2019,[47] while potentially smaller especially in the longer-run, remain positive and significant.

[47] 2006 is the start of the rapid rise in interest for renewable energy.

Figure D1: Renewable energy venture capital deals responses to EnvP-RE Shock, VAR fit to Monthly U.S. data under alternative specifications and samples



E Additional statistics and results: stock returns

Descriptive statistics - firm-level estimations and VAR

We consider the following 5 market risk factors: 1) MKTR is the market factor measured as the difference between the returns of diversified portfolios of the overall market and the safe interest rate at the end of month t, 2) SMB is the size factor measured as the difference between the returns of diversified portfolios consisting of stocks of small firms and big firms at the end of month t, 3) HML is the value factor measured as the difference between the returns of diversified portfolios comprising stocks of firms with a high book-to-market equity ratio and firms with a low book-to-market equity ratio at the end of month t, 4) RMW is the profitability factor measured as the difference between the returns of diversified portfolios consisting of stocks with robust and weak profitability at the end of month t, 5) CMA is the investment factor measured as the difference between the returns of diversified portfolios consisting of stocks with low (conservative) and high (aggressive) investment.

Industry classifications are defined according to ICB as follows: Basic Materials, Consumer Discretionary, Consumer Staples, Energy, Health Care, Industrials, Technology, Telecommunications and Utilities. We exclude observations with negative equity or sales values and observations where growth in total assets was larger than 100% in absolute value.

Table E1 shows summary statistics of all the variables used in the panel data analysis. In the estimations, environmental policy innovations and CO_2 emissions have been standardized to a mean of zero and a unit standard deviation and all financial variables are included in logs.

Table E1: Firm-level panel estimations, stock returns - summary statistics

	Observations	Mean	Std. Dev	Min	Max
Environmental Policy Innovations:					
EnvP innovations	164	1.8	25	-57	82
EnvP net sentiment innovations	164	0.0023	0.0181	-0.0422	0.0518
Policy Exposure Variables:					
AVG Scope-1 Emissions (mln tCO2)	49143	5	17	0.0005	128
AVG Scope-1 Emission Intensity (tCO2/ market cap)	49143	564	1462	0.32	11581
Pre-sample Scope-1 Emissions (mln tCO2)	34079	7	21	0.000006	150
Financial Variables:					
Excess returns	49143	-0.08	0.15	-0.88	0.60
Leverage (debt/equity)	49143	1.5	7.1	0	347
Firm size (mln market cap)	49143	26.3	56.4	0.05	1167.2
Profitability (return on assets)	49143	8.1	5.2	0.03	60.8
Dividends per share	49143	1.2	1.8	0.009	33.1

The environmental policy innovations are residuals extracted from an AR(7) and AR(6), respectively. All financial variables are GDP deflated. The sample excludes the recession period associated with the GFC.

Table E2 shows summary statistics of all the variables used in the VAR analysis.

Table E2: VAR models, ETF clean energy - summary statistics

	Observations	Mean	Std. Dev	Min	Max
Environmental Policy Index:					
Δ Log EnvP Index	149	0.00	0.32	-1.02	0.78
Economic Control Variables:					
Δ Log Oil (WTI)	149	0.00	0.08	-0.25	0.21
Δ Fed funds rate	149	0.03	0.08	-0.27	0.25
Δ Log NYSE Tech 100	149	0.01	0.03	-0.13	0.09
ETF Variables:					
Δ PBW ETF Assets under Management	149	4.07	38.94	-102.22	196.64

The sample excludes the recession period associated with the GFC as in our VAR.

VAR: robustness analysis

Figure E1 shows that our VAR results are robust to alternative Cholesky orderings.

Figure E1: Robustness results - Estimated effect of a shock in EnvP-RE news index on PBW-ETF market cap changes, monthly