Dissecting Climate Change Risks: Are they Reflected in Stock Prices?*

Renato Faccini[†‡] Rastin Matin[‡] George Skiadopoulos[§¶]

February 1, 2021

Abstract

We construct novel proxies of physical and transition climate change risks by conducting textual analysis of climate change news. The analysis of climate change news over 2000-2018 reveals that four textual variables are related to news on U.S. climate policy, international summits, natural disasters, and global warming, respectively. The first two topics proxy transition risks, whereas the last two topics proxy physical risks. We find that only the climate policy factor is priced in the U.S. stock market with the evidence being more pronounced over 2012-2018. The documented premium is consistent with the idea that investors hedge transition risks. We validate this explanation by constructing a narrative factor. Our results imply that investors' attention is an important driver of asset returns.

JEL classification: C63; G12; Q5

Keywords: Textual analysis; Latent Dirichlet Allocation; Physical and transition climate risks; Cross section of stock returns

*Emails: rmmf@nationalbanken.dk, rma@nationalbanken.dk, and g.skiadopoulos@qmul.ac.uk; gskiado@unipi.gr. We would like to thank Giorgio Mirone and Marcus Mølbak for their useful suggestions. We also thank seminar participants at Danmarks Nationalbank and at the conference on "Climate Change, Pandemics, Monetary Policy: New Approaches in Crisis Time" for their comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of Danmarks Nationalbank, or those of the European System of Central Banks.

†Centre for Macroeconomics (LSE)

‡Danmarks Nationalbank

§Department of Banking and Financial Management, University of Piraeus

¶School of Economics and Finance, Queen Mary, University of London

1 Introduction

The risks from climate change are physical (e.g., hurricanes, rise of sea levels, wildfires), and transition risks, emanating from the transition to a low carbon economy, which may impose costs to firms (e.g., carbon taxation, emergence of competitive green technologies). In this paper, we identify these two types of risk and examine whether they are reflected in U.S. stock prices. The answer is not obvious in advance, given investment practices and results from surveys, and it is of importance to both policymakers and investors.[1] If climate change risks are not priced by financial markets, policymakers should intervene and adjust their tax policies, emission limits, and cap-and-trade emission schemes.[2] This will affect the profitability and operation of firms, and as a result the returns of investors' portfolios.

Given the multifaceted nature of climate change risk, our investigation proceeds in two stages. In the first stage, we obtain novel climate textual factors which capture different sources of physical and transition climate change risks. In the second stage, we conduct the asset pricing tests, we propose an economic explanation for the results, and we validate it by a series of tests.

To construct our climate risk factors, we employ the Latent Dirichlet Allocation (LDA, Blei et al. (2003), Hansen et al. (2017)), an unsupervised textual analysis method. We apply LDA to the articles that contain the words "climate change" at least once, published over 1 January 2000 - 31 December 2018 in Thomson Reuters News Archive, a leading provider of information to the financial sector. LDA classifies the news corpus into categories (termed topics). Each topic contains a set of words ranked by the frequency they appear in the topic. Once the method delivers the categories, the user labels

[1]On the one hand, some institutional investors may not regard climate change risks as important as other financial risks (Krueger et al. (2020)). Institutional investors (e.g., pension funds) who advocate a "decarbonisation of portfolios" approach, according to which investments should be reallocated to green assets, face constraints in its practical implementation (Bessembinder (2017)). In addition, the effects of transition risks may not be easy to identify; these may depend on government's intervention, tastes of consumers, and the education of investors (CFA Institute (2020)). Finally, stronger investors' preferences for sustainability may result in a positive pricing effect for high sustainability stocks, thus causing a mispricing (Gibson et al. (2020)). On the other hand, climate change risks which are incorporated in legislation have imminent effects and may affect investors' decisions.

[2]Climate-related risks may threaten financial stability. Threats to financial stability arise when risks are not properly priced; the mispricing of mortgage backed securities during the first decade of the 2000s, played a key role in preparing the conditions that eventually led to the Great Recession.

each category based on the frequency and type of words being included. In addition to topics, LDA also delivers the topic shares, that is the probability that any given article is associated with any given topic. Given that articles are time-stamped, topic shares form a time series for any given topic.

Our corpus of articles is heterogeneous, encompassing various dimensions of climate change risk. It contains news ranging from the political debate on climate change legislation to news on natural disasters, the role of emissions on global warming, and corporate actions related to climate change. We single out four relevant topics which have a clear interpretation and they are potentially relevant to the stock market: the occurrence of natural disasters, the role of emissions on global warming, U.S. climate policy (actions and debate), and international climate-change summits. We treat the time series of their respective topic shares as climate risk factors. The interpretation of an increase in the value of the climate risk factor depends on the type of the climate risk factor under consideration. An increase in the natural disasters, and global warming factors would signify an adverse effect to the economy. This type of news attracts media's attention in the case where its content constitutes a source for concern for the society and the economy. An increase in the international summits factor also signifies an adverse effect to the economy; their main objective is to discuss the introduction of a global tax on pollutants, which is "bad news" for the economy in the short run. On the other hand, an increase in the coverage of the U.S. climate-related policy news may signal an increase or a decrease in transition risks, depending on which party in the U.S. holds the political power.[3]

Next, we employ the universe of U.S. common stocks over the same period and we investigate whether each climate factor is priced. To conduct our tests, we employ a standard portfolio sort approach. We sort stocks in value-weighted portfolios based on the sensitivity of each stock's returns to each factor (climate beta). Then, for any given

[3]In our sample, with the exception of the first term of Barack Obama's administration, the U.S. political debate on climate change has hardly ever pointed towards a likely increase in transition risks. Notably, the time period of our analysis has witnessed two climate-change denialists as presidents of U.S.A., George Bush and Donald Trump. Moreover, the second term of Obama's administration has not been successful in passing environmental friendly legislation through Congress, since the President lacked the required majority in the House of Representatives after the November 2012 elections and then also lost the majority in the Senate after the mid-term November 2014 elections. Therefore, over the examined period, one would expect that an increase in this factor would signify a positive effect to the economy, since it would reveal a reduction in short-term transition risks.

climate risk factor, we examine whether a long-short spread portfolio consisting of going long the portfolio which includes stocks with the greatest climate betas minus the portfolios, and short in the portfolio which includes stocks with the smallest climate beta earns a statistically significant average return, once we control for other risk factors. If it does, this would suggest that the climate change risk, proxied by the specific climate factor, is priced. To test the robustness of our results, we sort stocks in decile and quintile portfolios, separately, and we use alternative specifications to estimate stocks' climate betas and the long-short portfolios' alphas.

We find that only the U.S. climate policy factor is priced. The spread portfolio formed on the U.S. climate policy factor earns a statistically significant positive alpha, in almost all cases. In the case where we consider decile (quintile) portfolios, the spread's portfolio alpha ranges between 0.46% to 0.96% (0.30% to 0.59%) per month across models to estimate climate betas and alphas. There is no evidence that the risks elicited by news about the occurrence of natural disasters, the effects of emissions on global warming, and risks elicited from international summits are priced. We confirm the evidence that these three factors are not priced by performing our analysis over both monthly and annual investment horizons. Interestingly, the U.S. climate policy factor in not priced over the annual horizons. One explanation could be that investors may regard policy actions to be reversible because the views of politicians often change depending on the power game of political parties.

The existence of a positive risk premium for the U.S. climate policy factor is consistent with the interpretation that an increase in the value of this factor translates to good news for the economy (i.e. it decreases transition risks), and it can be explained by a hedging argument. A decrease in the U.S. climate policy factor, deteriorates the investor's opportunity set. To hedge against such an unfavourable shock, investors would hold (short sell) stocks with negative (positive) climate betas, thus increasing (decreasing) their prices and reducing (increasing) their return. As a result, the long-short portfolio (i.e. high climate betas stocks minus low climate beta stocks) would yield a positive alpha, as we find.[4]

[4]Interestingly, our obtained alphas are greater than these reported in Hsu et al. (2020) who also study the alpha of spread portfolios sorted on the firm's pollution intensity (firm's emissions divided by its assets) to assess whether environmental regulatory risks are priced in U.S. stocks. They find that the

We verify that the hedging argument explains our finding on U.S. climate policy being priced, by following two sequential steps. First, we find that the statistical significance of the climate policy factor hinges on the latest part of the sample, 6 November 2012 - 31 December 2019. Over this period, characterized by the second term of Obama's administration and the one of Donald Trump, news have typically signalled a reduction of transition risks.[5] Second, we carry out a narrative analysis to identify the content of climate change news (for an application in economics, see Romer and Romer (2004))[6]. We collect all articles which load with more than 40% on the topic. This yields 3,500 articles. We read each article, and mark it according to whether it captures an increase or a decrease in transition risks. By construction, an increase in this narrative factor reflects an increase in transition risks. We find that the narrative factor is priced and it carries a negative risk premium in the post-2012 period. This confirms the results from the analysis with the textual policy factor that risks stemming from the U.S. political debate are priced in the U.S. market because investors seek to hedge against it. Stocks which are more exposed to the textual (narrative) factor are riskier because a decrease (increase) in the factor signifies an increase in transition risks. To hedge the risk of the textual (narrative) factor, investors buy stocks with negative (positive) climate betas, thus increasing their prices and lowering their returns. As a result, the long-short spread portfolio formed with respect to the textual (narrative) factor will yield a positive (negative) alpha. The fact that U.S. climate policy is priced in a pronounced way in the post-2012 period is

CAPM alpha of the spread portfolio constructed using quintile portfolios is 4.07% per year over October 1992-September 2018. On the other hand, our CAPM alpha for the U.S. climate policy factor is 7.98% per year in the case of quintile portfolios considered over January 2000 - December 2018. Apart from differences in time periods, the difference may be attributed to the different sorting variables since the magnitude of textual betas is expected to be greater than that of pollution betas; data on pollution are updated once a year whereas our textual factors may vary significantly every day.

[5]The lack of a majority in the House of Representatives over Obama's second term in Office, and then also in the Senate after November 2014, forced the Democratic administration to come to terms with the Republicans in order to resolve the political impasse. As a price to pay, the Obama's administration eventually scaled down their ambition to tackle climate change. Trump continued to unravel any progress made by the Obama's administration on climate change issues (e.g., the appointment of Scott Pruitt, a notorious climate change denialist, as head of the Environment Protection Agency), ultimately withdrawing from the International Paris agreement. These news are "good" for the economy in the short run. The realization of transition risks entails a temporary negative impact on production, the price that needs to be paid to curb climate change.

[6]An alternative approach to decide on whether the content of news has a positive or negative meaning, would be to apply a sentiment correction using dictionary based methods. In the absence of climate dictionaries, we resort to a narrative approach to avoid the mis-classification of the content of news (LOUGHRAN and MCDONALD (2011))

also consistent with previous findings that climate change has risen to the attention of investors, only in the most recent years (Bolton and Kacperczyk (2020b), Painter (2020)).

Our findings suggest that only the immediate transition risks stemming from the domestic political debate on climate are priced whereas longer-term transition and physical risks do not appear to be priced. The four climate change topics elicit information on different sources of climate change risk, whose effects spread over different time horizons. News about the occurrence of natural disasters and global warming are informative for mostly long-run physical and transition risks. They reveal the direct effects of climate change on the current, as well as on future production because of rising temperature and the associated occurrence of extreme meteorological events.[7] On the other hand, articles about U.S. climate policy are informative about very short-term transition risks. These articles include news on the political debate on climate change, appointments in key positions in organisations like the Environment Protection Agency, and related laws passed in the Congress. Therefore, they reflect political intentions and actions over the course of the government's administration, i.e. at most four years; political positions in the Congress may radically change with a new round of elections. These views may well change, even if the same political party is re-elected, when there is a shift in the political composition of the Congress; the change in the environmental policy of Barack Obama's government in its second term is an example. Finally, articles about international summits are informative about transition risks over a longer time horizon relative to U.S. climate-policy news. This is because agreements decided in international summits take more time to filter through the climate policy debate, and eventually become law, if they ever do.

[7] According to the International Monetary Fund (2020) (IMF), under the current emission policies, global temperatures could increase by an additional 2–5°C by the end of this century, reaching levels not seen over the past millions of years. This would cause physical damage and an adverse effect to the economy. The latter would stem from greater temperatures, which implies a loss in output from agricultural crops and fishing farms, as well as reduced productivity for people working outdoors. It would also stem from physical damage to productive capital, infrastructure and buildings, due to the more frequent occurrence of natural disasters. Scientists have warned that temperature increases relative to pre-industrial levels need to be kept well below 2°C — and ideally 1.5°C — to avoid reaching climate tipping points and imposing severe stress on natural and socio-economic systems. According to the Intergovernmental Panel on Climate Change (2018) (IPCC), reducing the increase in temperature from 2°C to 1.5°C, could reduce the number of people susceptible to poverty due to climate-related risks by up to several hundred million worldwide. To this end, net carbon emissions need to decline to zero by mid-century (Intergovernmental Panel on Climate Change (2018)), which in turn requires prompt policy responses.

Our results are also consistent with the view that investors' attention is an important driver of asset returns. The approval of climate-related bills is a "wake-up" call for investors. This echoes the results in Choi et al. (2020), where global warming is noticed by retail stock investors in periods with unusually high temperatures.

Our paper is related to the growing literature on the measurement of climate change risks and their effects to asset prices. From a theoretical perspective, Hsu et al. (2020) show that the stocks of firms which pollute more (brown firms) than others (green firms) should command a greater risk premium because they are riskier; they are more exposed to penalties imposed by regulation, should the regulator decide to penalize them (environmental regulation uncertainty). Pastor et al. (2020) develop an asset pricing model with an environment-society-governance (ESG) factor which can also accommodate climate change risk. In their model, brown assets command a greater expected return than green assets because investors have green tastes and/or green assets hedge climate risk; our findings are consistent with their argument. Pedersen et al. (2020) present an ESG-adjusted capital asset pricing model, where ESG raises or lowers the required return, depending on the interaction of three groups of investors: those who do not take ESG into account, those who take ESG into account in forming expectations, and those whose preferences are also affected by ESG. Zerbib (2020) provides an asset pricing model which shows how the two common practices of sustainable investing (excluding sin stocks and taking ESG into account) affect expected stock returns.

From an empirical perspective, the literature uses different variables to proxy the risk stemming from climate change, and finds mixed results on whether a climate change risk premium exists, and on its sign. Hong et al. (2019) find that the increasing risk of droughts caused by global warming is not efficiently discounted by food stock prices. Baldauf et al. (2020) find little evidence that the flood risk due to sea level rise is priced in coastal real estate prices, whereas Asaf et al. (2019) find that this specific risk is priced. Painter (2020) finds that the underwriting fees and yields at issuance of long-term municipal bonds are affected by the risk of sea rising levels, whereas this is not the case for short-term municipal bonds. Ilhan et al. (2020) find that out-of-the-money options are relatively more expensive for carbon intense firms. Focusing on stock markets, the climate change risk premium, proxied by carbon risk (carbon risk premium) is found to be zero (Görgen et al. (2019)), positive (Bolton and Kacperczyk (2020b), Bolton and Kacperczyk (2020a), Hsu et al. (2020)), or negative (In et al. (2019)). Differences in results may be attributed to different periods, data sources, measurement of pollution, and construction of scores. Based on a survey of institutional investors , Krueger et al. (2020) underline that climate change risk may not be priced because climate risks are found to be difficult to price and hedge. Most closely related to our paper is Engle et al. (2020), who also use textual analysis to construct their aggregate climate change risk measure. We differ from their paper in two ways. First, they do not distinguish between different types of climate change news. As they note: "*Separately measuring news series about physical and regulatory climate risk represents an interesting avenue for future research.*" Second, they do not proceed in testing whether their measures are priced in the cross-section of U.S. equities.[8]

Our results reconcile some of the above seemingly different findings reported in the literature. The results of Bolton and Kacperczyk (2020a), Bolton and Kacperczyk (2020b), Hsu et al. (2020), and Ilhan et al. (2020) who find evidence that climate policy uncertainty related to the treatment of carbon emissions is priced in the stock and option markets, are consistent with our finding that the U.S. stock market reacts to the news on the U.S. political debate on climate change; these load very heavily on topics related to energy production and emissions. Given that our climate policy textual factor includes news on *discussed*, as well as on *approved* policy plans, our results show that it is *both* climate policy uncertainty and approved climate policy plans that affect stock prices. Similarly, the evidence in Hong et al. (2019) and Baldauf et al. (2020), on food stock prices being unrelated to the increasing risk of droughts, and house prices being unrelated to the risk of sea level rise, are also consistent with our findings that stock market prices do not reflect longer-term physical risks.

[8] Li et al. (2020) and Sautner et al. (2020) employ textual analysis on conference calls of publicly-listed firms to construct measures of exposure to climate change at a *firm*-level. Their focus is on the relation of their measures with firms' characteristics rather than whether they are priced. Interestingly, our identified topics echo the natural and regulatory topics obtained by a similar in spirit to LDA textual method in Sautner et al. (2020).

2 Data and textual analysis

2.1 News articles from Reuters

Our sample consists of more than 13 million articles from Thomson Reuters News Archive published in the period from January 1, 2000 to December 31, 2018. Reuters News reaches one billion individuals each day, and its associated trading platform Eikon has a 34% market share for the delivery of financial information.[9] Reuters is thus a key player in this market, thus affecting stock market prices via the dissemination of news.

We restrict the analysis to news articles written in English and we apply filters to remove entries that summarize different unrelated news, or simply report tables of stock market returns. If there are subsequent corrections to an article, we use the first version of the article within a 12-hour period, and in case of additions to an article within a trading day, we use the article with the longest body text.[10] After this initial procedure, we end up with a sample consisting of roughly 7 million articles. This sample contains articles within a diverse set of topics, including sports, technology, politics, finance, among others. Given our focus on climate risk, we discard irrelevant articles by retaining only the news in which the biagrams "*climate change*" or "*global warming*" occur at least once. This yields a final sample consisting of roughly 34 000 articles.[11]

This textual corpus comprises a very heterogeneous set of articles related to climate change. Some articles reflect climate change views expressed in the domestic political debate over different geographical locations in U.S. and internationally; others reflect corporate views or marketing initiatives across the globe related to climate change; others report news about scientific research and on the effects of emissions on global warming; some news may report on the realizations of extreme meteorological events; finally, some news may be only incidentally related to climate change. To group the heterogeneous news into specific climate-subcategories, we conduct textual analysis by employing the

[9] https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/reuters-news-agency/fact-sheet/reuters-fact-sheet.pdf

[10] As soon as a news item occurs, Reuters publishes immediately a breaking news alert, often consisting of a single sentence. The body of the article is then added within a few minutes. In our corpus, we observe both entries separately, but we use only the second, updated version in the analysis.

[11] We found that this simple sub-selection was adequate due to the semantically independent topics subsequently estimated by our topic model. In general, our analysis is not sensitive to a minor residual of irrelevant articles as they will be labelled by the topic model and manually discarded.

Latent Dirichlet Allocation (LDA). We describe the method in the following section.

2.2 Latent Dirichlet Allocation: Concepts and estimation

LDA (Blei et al. (2003)) is one of the most commonly employed topic models in textual analysis (Zhao et al., 2015); it was first used in the economics literature by Hansen et al. (2017). It is a textual method which takes a collection of articles and the number of unique words (termed vocabulary) contained in these articles, as inputs. It delivers two outputs: first, the entire textual corpus is broken into categories (termed topics). Second, every article is expressed as a vector of weights over topics (termed topic shares). A topic is a probability distribution over the unique words: it reflects how frequently every unique word appears in a topic. This enables the user to label the delivered topic, based on the words which appear most frequently. LDA also allows us to retrieve how the news coverage of a given topic varies over time, by tracing the evolution of a given topic share. Because articles are time-stamped, time variation in a topic's coverage can be retrieved by summing, for every point in time, the weight of any given topic across articles.The number of topics is set by the user. In our case, we have $33,735$ articles and a vocabulary of $6,158$ unique words that appear across all articles.

LDA is a natural choice for the purposes of our analysis because it can deliver climate risk factors which capture different dimensions of climate risk and have a clear interpretation. LDA is an unsupervised machine learning method, i.e., it dissects textual heterogeneity in topics delivered by the method, rather than by the user. In contrast to dictionary methods, where the user labels the topic in advance by specifying the words that are most likely to characterize it, LDA yields the topics, and then the user labels them based on the words which are most frequently appearing in the topic. This is beneficial for our purposes because in the context of climate change news, words like "pollution", could feature in articles covering different themes, ranging from scientific research and corporate announcements, to natural disasters and climate-change legislation. So in practice, making assumptions on the words that characterize topics is a non-trivial task.

To fix ideas, LDA is a Bayesian factor model for discrete data. In a model with K topics, each topic is a probability vector β_K, over the Y unique words in the textual corpus. LDA is a mix-membership model, in that each article can be associated with multiple topics. Each article loads on the different topics with a vector of factor loadings, termed topic shares, θ_K. For a given number of topics K, both the topics (β_K) and the topic shares (θ_K) are outputs of the method. Once the model is estimated, a given article is represented as a vector of loadings over the K topics.

The probability that any given word in article x is equal to the y^{th} word in the vector of unique words is $p_{x,y} = \sum_k \beta_k^y \theta_x^k$. Let $n_{x,y}$ denote the number of times that word y appears in article x. Then, the likelihood is given by $\prod_x \prod_y p_{x,y}^{n_{x,y}}$. LDA assumes Dirichlet priors of the two probability distributions, i.e. priors of the probabilities of unique words belonging in topic k (vector β_K, $k = 1, ..., K$) and the probabilities of articles belonging to topic k (i.e. the topic shares in vector θ_K). The prior distribution for β_K is assigned a symmetric Dirichlet prior with Y dimensions and hyperparameter α. The prior distribution for θ_K is assigned a symmetric Dirichlet prior with K dimensions and hyperparameter η. The hyperparameters measure the concentration of the realizations. A high value indicates that the distributions are relatively flatter, with a relatively even distribution of the probability mass. In line with Heinrich (2009), we set $\alpha = 1/K$ and $\eta = 1/10$

We apply the Kalman filter to evaluate the likelihood function. Then, we obtain the posterior distribution by using the estimated likelihood function and the prior distributions for β_K and θ_K. The posterior kernel is simulated numerically using the Metropolis-Hasting algorithm. We relegate the technical details to the Appendix. LDA will then deliver one posterior distribution for β_K used to label the topics, and $33,735$ posteriors for θ_K for the respective $33,735$ articles (i.e. a matrix 33,735 by 25 of posterior probabilities) which will be the topic shares.

We select the number K of topics, so that the estimated posterior distribution for β_K yields topics which can be interpreted by the user. To this end, the most frequently encountered words within a given topic should be semantically similar. Our choice of $K = 25$ topics achieves this and yields a set of topics that are semantically independent.

2.3 Estimated topics: Interpretation

Within the corpus of climate change articles, our LDA model classifies the unique words in 25 different topics. We use the heat map reported in Figure 1 to interpret these topics.

For every topic, we order first the most frequent word, and then words follow in decreasing order of frequency. We use darker (brighter) colors for words with higher (lower) relative frequencies.

[Figure 1 about here.]

We can see that Topic 1 is about scientific research documenting how marine life became endangered as a result of global warming. A few topics are related to climate policy discussions about climate change taking place in different countries: topics 2 and 6 relate to Germany, topic 3 to Canada, topic 5 to Australia, topic 15 to a mix of countries including Africa, Indonesia and Brazil, topic 19 to Asia, topic 21 to the UK, topic 22 to Russia and Norway, and topics 4 and 7 focus on the U.S. Topic 4 primarily focuses on U.S. energy policy and its connections with the climate change debate at the State level, whereas topic 7 seems to be closely related to the debate of U.S. climate policy at the Federal level.

Topic 10 reflects news on renewable energies, with a focus on solar and wind technologies, as alternatives to more polluting energy sources like coal. Topic 17 relates to news about the effects of fossil-fuel emissions on global warming. Topic 18 collects news on international summits, where political leaders of many countries meet to discuss issues related to climate change, in an attempt to reduce global emissions. This topic includes news surrounding international events. Examples include the United Nations Copenhagen Conference of 2009, where representatives from 115 different countries met, as well as news that relate to discussions about the Kyoto protocol of 1997, an international treaty involving 192 parties, where nations agreed to reduce greenhouse emissions. Topic 14 also reflects news about the implementation of the Kyoto protocol, with a particular focus on the decisions taken at the level of the European commission.

Topic 19 is about political activism around climate change issues, whereas topic 25 seems to be related to news about the oil market. Topics 8, 20 and 23 broadly reflects corporate news. The remaining topics, 9, 11, 12 and 13 do not seem to reflect a clear theme, or one that can be clearly associated to a specific aspect of climate news.

For the purpose of our analysis, we will use four topics which have a clear interpretation and which are expected to be relevant to investors interested in U.S. equities: U.S.

climate policy (the union of topics 4 and 7), international summits (topic 18), natural disasters (topic 24) and global warming (topic 17). We discard from our analysis the topics related to climate policy legislation in all countries other than the U.S.; U.S. data are the benchmark in the asset pricing literature, and keeping the focus on this market facilitates comparability of results. We also discard topics that relate to corporate news since they tend to carry company-level information rather than information about a specific source of climate risk. For the same reasons we leave out information that relates to renewable energies. Moreover, both corporate and renewable energy factors reflect news that is not restricted to the U.S. market. Finally, we also discard the topic about maritime life research, oil and political activism, since the scope of these topics seems to be very narrow compared to that of natural disasters and global warming.

3 Topics as risk factors

As we have discussed, LDA delivers two outputs: topics and topic shares. We use the latter output as a risk factor. On any given day, LDA relates each article to each one of the topics in Figure 1, with some estimated weight. This weight is the topic share, i.e. the percentage of words in the article associated with this topic. For example, the estimation of the LDA model may yield that a specific article is 70% about topic 18, 20% about topic 24, and 10% about topic 17. Thus, these weights are probabilities and they add up to one, for any given article. For any given topic, the topic shares form a time series, given that articles are time-stamped.

On any given day, we may have more than one article. For each one of our four topics ($k = 1, 2, 3, 4$), we create a time series of news coverage by summing, for each day, the topic shares θ_k of every article related to climate change. We identify four risk factors which have a clear interpretation and which may potentially be relevant to investors interested in U.S. equities. Table 1 shows the pairwise correlations between these four factors (U.S. climate policy, international summits, natural disasters and global warming), as well as their correlations with standard equity factors used in the asset prcing literature (market fator, the value and size Fama-French factors (Fama and French (1993)), the momentum factor (Carhart (1997)) and the investment and profitability Fama-French factors (Fama and French (2015))). We can see that the pairwise correlations between the climate textual factors are small and not greater than 0.3. In addition, the correlations of the climate textual factors with the equity factors are also small.

[Table 1 about here.]

The low pairwise correlations between the four textual factors imply that these time series convey different information regarding climate risk. This comes as no surprise, as it is reasonable to expect that different topics elicit information on different sources of climate change risks. News about the protracted effect of emissions on global warming and about the occurrence of natural disasters carry information about long-run physical and transition risks: rising temperatures increase the probability of extreme meteorological events in the future. In turn, this calls for policymakers to tax polluters, and for firms to switch to more environmentally friendly technologies. Climate change scientists agree that the climate is deteriorating fast. This will force firms and policymakers to act accordingly in the next one or two decades. Changes in technology and consumer preferences, as well as increasing taxation, pose a cost to businesses and thus a threat for financial stability.

On the other hand, articles about U.S. climate policy are informative about very short-term transition risks. These articles include news on the political debate on climate change, appointments in key positions in organisations like the Environment Protection Agency, and related laws passed in the Congress. They represent short-run risks because they reflect political intentions and actions over the course of the government's administration, i.e. at most four years; political positions in the Congress may radically change with a new round of elections. These positions may well change, even if the same president is re-elected, when there is a change in the political composition of the Congress; the change in the environmental policy of Barack Obama's government in its second term is an example. Relative to these domestic climate-policy news, articles about international summits are instead informative about transition risks over a longer time horizon. This is because agreements decided in international summits take more time to filter through the climate policy debate, and eventually become law, if they ever do.

We interpret climate news as a risk factor because the disclosure of news reveals physical and transition risks. The interpretation of the meaning of an increase in the climate risk factor depends on the type of the climate risk factor under scrutiny. Typically, an increase in the natural disasters and global warming factors would signify an adverse effect to the economy. This is because this type of news attracts media's attention in the case where its content constitutes a source for concern for the society and the economy. An increase in the international summits factor also signifies an adverse effect to the economy; the main objective of these meetings is to discuss the introduction of a global tax on pollutants, which is "bad news" for the economy in the short run.

On the other hand, it is not clear in advance how one should interpret an increase in the U.S. climate policy factor. An increase in the coverage of this type of news may signal an increase or a decrease in transition risks, depending on which party in U.S. holds the political power. In our sample, with the exception of the first term of the Obama's administration, the U.S. political debate on climate change has hardly ever pointed towards a likely increase in transition risks. Notably, the time period of our analysis has witnessed two climate-change denialists as presidents of the U.S.A., George Bush and Donald Trump. Moreover, the second term of the Obama's administration has been characterized by the failure to pass any significant legislation through Congress, since the president lacked the required majority in the House throughout his second mandate and also in the Senate after 2014. Following the elections of September 2012, it became fairly clear that any effort to tackle climate change was unlikely to be effective, and that many of the ambitions of the Obama's administration would be scaled down.

Next, we delve in the content of news reflected in each of the four risk factors and on the specific press releases that make these factors vary over time. Figures 2a, 2b, 2c, and 2d show the time series evolution of the four respective risk factors; we depict the monthly average over the daily values for each month. The factors reach their highest values in 2007. This is due to an increased coverage of important climate-related news in 2007, as we describe below. It is likely that the media's attention toward climate change was also attracted by the award of the Peace Nobel Price to Al Gore and the Intergovernmental Panel on Climate Change (IPCC) in that year for "their efforts to build up and disseminate greater knowledge about man-made climate change, and to lay the foundations for the measures that are needed to counteract such change". We comment on the various peaks of each series by tracing the news that corresponds to them.

Figure 2a shows the time series of the natural disasters textual factor. This directly reflects news on the occurrence of catastrophic natural events, including the record highs of rainfall and drought in Asia in November 2000, the extremely cold winter in Europe in January 2006, Hurricane Dean in August 2007, flooding in Eastern India in August 2008, wildfires in Australia in February 2009, Cyclone Pam in March 2015, extreme pollution in New Delhi in November 2015, and wildfires in California in November 2018. The factor also reflects the content of scientific research and government reports that emphasize the role of climate change for the occurrence of natural disasters. Examples include the report by the Asia Development Bank in February 2012, which warned about the risk of mass migration due to the increased occurrence of natural disasters in the region, and the third United Nations (U.N.) conference on Disaster Risk Reduction in March 2015.

Figure 2b plots the time series of the global warming factor. This reflects mostly news on the effects of emissions on global warming. This news appears in multiple sources, including reports drafted by governmental and non-governmental organizations, both at a national and international level, publication of scientific studies in academic journals, and articles appearing in non-scientific magazines. This may explain the heterogeneity of this type of news which makes articles to have smaller weights (topic shares) on the global warming topic, relative to the natural disasters topic. As a result, the global warming factor can be related less often to a significant event, relative to the case of natural disasters. Examples where a strong association can be established, include the cases where the IPCC (February 2007, April 2007, November 2007), U.N. Panel on Climate Change (December 2009) and World Meteorological Organization (November 2015) reports are published. All these warned about the impacts of global warming and stress the need to reduce greenhouse gas emissions.

Figure 2c plots the time series of the international summits factor. This reflects the occurrence of international events, where governments' representatives from around the world meet to negotiate a coordinated intervention to tackle climate change. It also captures how legislation at a country level responds to these events. Indicative examples where our factor spikes to reflect the increased intensity of news on international summits include Hague talks (November 2000) and Bonn meetings (July 2001) which led to the ratification of the Kyoto protocol of 1997 (February 2005), the G8+5 meeting (February 2007), the Bali and U.N. Poznan and Bonn meetings (December 2007, December 2008, June 2009, respectively), the Copenhagen Summit (December 2009), and Doha U.N. climate change conference (November 2012), as well as amendments in the legislation such as the coordination of U.S. and European exchanges on emission trading schemes (May 2006). Following November 2012, the International Summits textual factor stays at a relatively low level. Interestingly, there is no pronounced movement in December 2015, when the Paris Agreement took place. This is because reference to it appears in the textual sample that follows that date. Under the Paris Agreement, it was established that each country must determine, plan, and regularly report on the actions that it undertakes to mitigate global warming. On the other hand, no mechanism forces a country to set a specific emissions target by a specific date.

Figure 2d plots the time series of the U.S. climate policy factor. The series reflects news releases on Presidential speeches, the outcome of elections in the House of Representatives and Senate with respect to their climate related implications, the discussion and introduction of environmental bills, the political consequences of natural disasters, and the appointment in key positions of people with well declared views on environmental issues. Examples include the George Bush and Barack Obama's State of the Union Address (January 2007 and February 2013, respectively), the Democratic and Republican parties taking control of the House of Representatives (November 2006 and November 2010, respectively), the passed bills on capping Greenhouse gas emissions for the first time and promoting the use of clean energy resources (June 2007, September 2009), the Lieberman-Warner Climate Security Act (June 2008), the introduced bills to stop the regulation of emissions and to approve the keystone XL pipeline (March 2011 and November 2014, respectively), the political aftermath of the BP oil spill in the Gulf of Mexico (April 2010), and the appointment of Scott Pruitt by Donald Trump to lead the Environment Protection Agency (December 2016).

Appendix B provides a detailed description of the news releases associated with some of the pronounced increases in the value of each one of our textual factor, including the above mentioned ones.

[Figure 2 about here.]

4 Asset pricing tests

We investigate whether each climate factor is priced in the cross-section of U.S. stocks. Our sample is unbalanced and spans the same period over which we have collected news, January 1, 2000 - December 31, 2018. We obtain daily stock prices from the Center for Research in Security Prices (CRSP). Our stock universe consists of all U.S. common stocks trading in NYSE, NASDAQ, and AMEX (CRSP share codes 10 and 11). For each day, we have on average about 4,700 returns, from a total of 10,498 listed firms in our sample. We adjust returns for delisting as in Shumway (1997). We also collect yearly data on the environmental pillar of the Thomson Reuters ESG scores.

4.1 Portfolio sort analysis

To conduct our asset pricing test, we employ a standard portfolio sort approach. We sort stocks into portfolios based on their sensitivity to each factor (climate beta). Then, we form a long-short spread portfolio consisting of going long the portfolio which includes stocks with the higher climate beta minus the portfolios, and going short in the portfolio which includes stocks with the smallest climate beta. We examine whether the spread portfolio earns a statistically significant abnormal performance If it does, this would suggest that the climate change risk proxied by the specific climate factor is priced.

To fix ideas, for every asset i, we estimate:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i F_t + \gamma_i X_t + \epsilon_t, \tag{1}$$

where $r_{i,t}$ is the daily return on security i, $r_{f,t}$ is risk-free return, F_t is the textual factor, X_t is a vector that includes standard controls that have been found to explain the cross section of U.S. stock returns and ϵ_t is an i.i.d. error term with zero mean. At the end of every month, we estimate equation (1) recursively, using a rolling window consisting of daily observations over the previous three months. We roll forward the starting date of the window by one month at each iteration. At the end of any given month, given the estimated betas across stocks, we rank stocks according to their estimated betas and group them in portfolios; we form decile and quintile portfolios, separately. Then, for each portfolio, we compute the portfolio's post-ranking value-weighted monthly returns. Next, we compute the long-short spread portfolio's monthly return. We repeat the process until we exhaust our sample. This yields a time series of 225 spread portfolio monthly returns. Finally, we estimate its alpha, and we assess its statistical significance. To estimate the spread portfolio's alpha, we use the same asset pricing model (i.e. the same set of factors X_t) as the one we employed in equation (1) to estimate the stocks' betas.

Table 2 reports the results on the estimated alphas (unit is % per month) and their $t-$statistics within parentheses. We report results for each one of our four climate change factors for the decile and quintile portfolio sorting, separately, and across five model specifications, regarding the choice of vector X_t in equation (1): the market model, which only includes the market portfolio return (market factor); the Fama-French three factor model (FF3, Fama and French (1993)), which controls for the market factor, as well as the size and book-to-market factors; the Fama-French-Carhart (FFC, Carhart (1997)) four factor model, which controls for the same factors as FF3, and also includes Carhart's momentum factor (umd); the Fama-French five factor model (FF5, Fama and French (2015)), which controls for the same factors as FF3, as well as for the profitability and investment factors; a specification that includes the momentum factor (umd) in addition to the factors included in FF5.

[Table 2 about here.]

We can see that the alphas of the long-short portfolios formed on the global warming factor are negative, yet statistically insignificant in all cases. In the case of international summits, we find that alphas are negative, yet statistically significant only for some specifications. In the case of natural disasters, alphas are positive in almost all cases, yet statistically insignificant. Therefore, we cannot reject the null hypothesis that the risks elicited by these factors are not priced. On the other hand, the alpha of the long-short portfolio formed on the U.S. climate policy factor is positive and statistically significant, in all but in the FF3 and FF4 specifications that rely on quintile sorting. In the case where we consider decile (quintile) portfolios, the spread's portfolio alpha ranges between 0.46% to 0.96% (0.30% to 0.59%) per month across models to estimate climate betas and alphas.

The statistically significant positive risk premium of the U.S. climate policy factor can be explained through the lens of an intertemporal hedging motive. An increase in the U.S. climate policy factor signifies "good news" for the economy over most of our sample, as we explained; in general, over our sample, the political declarations and actions decrease the transition risks. This would mean that a decrease in the U.S. climate policy factor, deteriorates the investor's opportunity set and future consumption. To hedge against this unfavourable shock, investors would hold (short sell) stocks with negative (positive) climate betas, thus increasing (decreasing) their prices and reducing (increasing) their return. As a result, the long-short portfolio (i.e. high betas stocks minus low beta stocks) would yield a positive alpha, as we find.[12]

To provide further support to our explanation for the existence of a positive risk premium for the U.S. climate policy factor, we carry out a sub-sample analysis. We take November 6, 2012, as a splitting point. This splitting point marks the beginning of the second term of Obama's administration and its choice can be justified as follows. According to our rationale stated above, the evidence on U.S. climate policy factor being a priced factor should be more pronounced from November 6, 2012 onward because post-November 2012 news signal inability, or reluctance, to tackle climate change. Hence, an increase in this factor is "good news" for the economy and thus it minimizes transition risks.[13] We test whether the U.S. climate policy climate factor is priced by repeating

[12]Interestingly, this hedging argument can also explain the negative, albeit insignificant, risk premiums of the global warming and international summits factors. As explained, an increase in the global warming, and international summits factors signifies an increase in physical and transition risks, and thus "bad news" for the economy, and a deterioration in future investment and consumption opportunities. To hedge against this unfavourable shock, investors buy (short sell) stocks whose returns increase (decrease) in times that these factors increase. This implies that investors buy (short sell) stocks with higher covariance with the factor. As a result, they pay higher (lower) prices and accept lower (higher) returns for stocks with higher (lower) climate factor betas. In this case, the long-short portfolio (i.e. high betas stocks minus low beta stocks) will yield a negative alpha, in line with our evidence.

[13]The period that follows the November 2012 elections was characterized by a a lack of a majority for Democrats in the U.S. House of Representatives. After 2014, Democrats also lost control of the Senate. As a result, President Obama was unable to fulfil his electoral promises, backtracking on the progress that was made during his first mandate. Climate change policy was one of the matters on which the Obama's administration eventually decided to abandon in search for a political compromise that would solve the political impasse. In addition, the period of Trump's administration has been characterized by a manifested aversion to tackle climate change. The President repeatedly declared that climate change is a hoax. The people who Trump appointed to lead the Environment Protection Agency, showcased his intentions. His first nominee, Scott Pruitt, was a notorious climate change denial. His second nominee, Andrew Wheeler, was notoriously associated with the coal lobby. Ultimately, Trump reneged on the Paris agreement, reversing any progress that Obama was able to make in his first mandate.

the portfolio sort analysis over the sub-periods January 1, 2000 - November 5, 2012 and November 6, 2012 - December 31, 2018. Table 3 presents the results. We can see that the alphas of the spread portfolio sorted on the U.S. climate policy factor are positive and statistically significant in the post-2012 period. In contrast, alphas are insignificant in the corresponding pre-2012 cases. Therefore, the U.S. climate policy factor is priced only in the second sub-period. Notably, after November 6, 2012, the U.S. climate policy factor is priced for both portfolio sorting schemes (decile and quintile portfolios), and across all model specifications, including those that showed lack of robustness over the full-sample. Moreover, in most of the cases, $t-$ statistics are close and even exceed the threshold of three suggested by Harvey (2017) to address data mining concerns (see also Hou et al. (2020)). These findings indicate that the evidence on U.S. climate policy being priced reported in Table 2 is driven by the period that follows the second Obama's mandate. Interestingly, the international summits factor appears to be priced only a some cases in the pre-2012 period, yet any significance vanishes in the post-2012 period. This finding is consistent with the fact that U.S. withdrew from agreements associated with international summits (e.g., Paris agreement) over that period, and this this factor posed no threat to polluters to require a compensation for being exposed to it.

[Table 3 about here.]

[Table 4 about here.]

To explore more the economics behind the evidence on climate related U.S. climate policy being priced in the U.S. stock market, we report characteristics of the quintile portfolios constructed by sorting stocks on the climate beta with respect to this factor over November 6, 2012 - December 31, 2018. In particular, we are interested in exploring whether there is any correspondence between estimated climate betas and environmental indicators. For this purpose, we rely on the environmental pillar indicator of the ESG scores produced by Thomson Reuters. It is well known that such indices are noisy, in the sense that their values differ substantially among different providers (Berg et al. (2020)). To minimize the effects of noise, we report results based on quintile sorting.

Table 4 reports the average value-weighted return, average climate beta for each textual factor, average value of the environmental pillar indicator of the ESG scores produced by Thomson Reuters, the percentage change in this ESG score, the average market capitalization and the average number of firms for each portfolio. For each textual factor, we estimate climate betas for the various model specifications reported in Table 2 and Table 3.

We can see that the firms that are most exposed to the risks elicited by U.S. climate policy (grouped in quintile 5), also tend to have a relatively lower ranking in terms of environmental performance. However, a seemingly puzzling feature arises: firms that are least exposed to the same risks (grouped in quintile 1), also tend to perform poorly in terms of environmental classification. Notably though, firms that are sorted in the first quintile tend to be those which have experienced the highest improvement in their environmental score. This pattern prevails regardless of the model employed to estimate climate betas and it is consistent with our explanation of the positive sign of the U.S. climate policy climate-related factor. It implies that investors hedge their climate change risk by investing in firms which show a strong intention to become environmentally friendly, even if the level of their current environmental score may be still low.

Our findings have two important implications. First, investors have only started taking climate risk into account recently. Our results suggest that any pricing of climate risks is a recent phenomenon, associated with the most recent years, covering less than a decade. Further breakdowns of the sample over the period before 2012, for instance, isolating the time of Obama's first mandate, also reveal lack of significance. These results are in line with the findings in Bolton and Kacperczyk (2020b) and Krueger et al. (2020), who also conclude that the pricing of climate risk is a recent phenomenon. Second, our results suggest that investors price only specific aspects of climate risk. In particular, they price these risks only when they come to the political arena, i.e., they are concerned only about short-term transition risks. On the other hand, investors do not take into account longer-term physical and transition risks captured by natural disasters, global warming and international summits. This is in line with the results of the survey conducted by Krueger et al. (2020), where the average respondent believes that equity valuations do not fully reflect the risks from climate change. These results are also consistent with the view that investors' attention is an important driver of asset returns. The approval of climate-related bills is a "wake-up" call for investors. This echoes the results in Choi et al. (2020), where global warming is noticed by retail stock investors in periods with unusually high temperatures.

Moreover, our findings highlight the advantage of taking a textual factor approach to explore whether climate change risk is priced. The advantage is twofold. First, the fact that we decompose climate change risk in its different aspects (physical and transition) allows us to reconcile some seemingly different findings reported in the literature. Bolton and Kacperczyk (2020b) and Bolton and Kacperczyk (2020a) find that transition risks related to carbon emissions are priced, at least in part. Hsu et al. (2020) reach similar conclusions, finding evidence of a pollution premium. These results are consistent with our finding that the U.S. stock market reacts to the news on the U.S. political debate on climate change, which load very heavily on topics related to energy production and emissions. On the other hand, Hong et al. (2019) find that increasing risk of droughts caused by global warming are not efficiently discounted by prices of food shares. Baldauf et al. (2020) find weak evidence of real estate prices falling in response to greater flood risk as the sea level rises. These results are consistent with our findings that stock market prices do not reflect longer-term physical or transition risks.

Second, in the case where one would consider an aggregate climate textual factor, this could mask important information for pricing purposes. We explore this by repeating the portfolio sort analysis using an aggregate textual factor constructed by counting the articles featuring the words "climate change" at any day. Table 5 reports the results. We can see that the aggregate textual factor is not priced, thus hiding the valuable information contained in news related to U.S. climate policy for the purposes of pricing the cross-section of U.S. equities. This confirms the necessity to decompose climate risk in its various aspects, and highlights the benefits of LDA as a textual analysis technique to address our research question.

[Table 5 about here.]

4.2 Fama-MacBeth regressions

The portfolio sorts analysis provides evidence that the U.S. climate policy is priced in the cross-section of individual U.S. equities. In addition, it is the 2012-2018 period that drives this evidence. We perform a further robustness test by conducting Fama-MacBeth (FM, Fama and MacBeth (1973)) regressions over the 2012-2018 period. FM regressions have the advantage over the portfolio sorts analysis that they can account for the effects of multiple regressors. On the other hand, they can only account for linear relations, whereas portfolios sorts can account for non-linear relations, too. As a result, the two approaches may not yield similar results. We perform FM regressions by examining two different cases: one that uses each one of the textual factors separately as a regressor, and one which uses all four textual climate factors simultaneously. We use as control variables the set of factors which appear in the various specifications we have used in the portfolio-sorts analysis to estimate climate betas. To minimise the effects of errors-in-variables, we use portfolios as test assets. We opt for a wide set of test assets: we use 55 and 74 industry portfolios separately. Both sets of test assets include the 25 Fama-French portfolios sorted on size and book-to-market. They differ in that the first also includes the 30 Fama-French industry portfolios, and the second includes the 49 Fama-French industry portfolios. The finer partition of stocks in industry sectors may reveal differences on how climate change risk may affect different industries. In the first-pass regressions, for each security, we estimate climate betas using a rolling window of the daily observations over the past three months. We repeat the procedure by rolling the beta estimation window by one month, just as we did in the portfolio-sort approach to asset pricing tests. In the second pass regressions, at each time step, we obtain the price of risk of each factor by running cross-sectional regressions of the stock returns over the next month on the estimated betas of the factors obtained from the first-pass regressions.

Table 6 reports the average over time price of risk of each factor and its t-statistic for the two different sets of test assets for the case where we use the Carhart (1997) specification to estimate betas. The expected return-beta representation equation is

$$\begin{aligned} E(r_i) - r_f = \lambda_0 + \lambda_{MKT}\beta^i_{MKT} + \lambda_{HML}\beta^i_{HML} + \lambda_{SMB}\beta^i_{SMB} + \lambda_{UMD}\beta^i_{UMD} \\ + \lambda_{ND}\beta^i_{ND} + \lambda_{GW}\beta^i_{GW} + \lambda_{IS}\beta^i_{IS} + \lambda_{CP}\beta^i_{CP} \end{aligned} \tag{2}$$

We can see that the U.S. climate policy factor is priced in most of the specifications for the set of control variables in the 2012-2018 period and the price of risk is positive.

This holds irrespectively of whether one uses the factor in a stand-alone fashion in the FM regressions (column (iv)), or jointly with the other climate textual factors (column (v)). It also holds regardless of whether one employs the 55 or 74 portfolios, thus implying that the U.S. climate policy does not affect differently firms that belong in the same sector of the economy. The other three climate textual factors are insignificant when considered in a stand-alone fashion or simultaneously (columns (i)-(iii) and column (v)). Therefore, the FM regressions confirm the results from the portfolio sorts analysis, that is, the climate policy factor is priced whereas natural disasters, global warming and international summits are not. Results are similar in the cases where the alternative specifications discussed in Section 4.1 are used in the first step FM regressions to estimate climate betas, and they are not reported due to space limitations. The control variables are not priced. This is in line with previous empirical evidence when relatively short periods are examined (Chang et al. (2013)).

[Table 6 about here.]

4.3 Portfolio sorts: Evidence from annual returns

Our results in Section 4 could be driven by the fact that we conducted our asset pricing tests by employing monthly post-ranking returns. In this Section, we explore whether this may be the case by performing our tests using annual post-ranking returns.

At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to a given textual factor (global warming, natural disasters, international summits and U.S. climate policy textual factors). Then, we compute the post-ranking value-weighted portfolio annual return, and the resulting spread portfolio return is computed as the difference between the return of portfolio 10 (high climate beta) minus the return of portfolio 1 (low climate beta). A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. We estimate the $averagereturnand$ alpha of the spread portfolio, from annual overlapping post-ranking returns, over the period January 1, 2000 - December 31, 2018.

Table 7 reports the results. We can see that the three textual factors (natural disasters, global warming and international summits factors) are found not to be priced, just as it was the case with the results obtained from the analysis based on monthly post-ranking returns. On the other hand, we find that the U.S. climate policy factor is not priced any longer, once we switch from monthly to annual horizons. This evidence corroborates our conjecture that the three former factors may represent long-run risks. Moreover, it corroborates our explanation that investors are concerned about short-run risks, as these are represented by the U.S. climate policy factor, and this explains why this factor is not priced over longer horizons. Investors may regard policy actions to be reversible because the views often politicians often change depending on the power game of political parties. Our evidence provides a fresh look to the results of Bansal et al. (2017) who find that long-run temperature fluctuations are priced in equity markets. The difference in results may be attributed to the different proxies for climate change risks (temperature versus a global warming textual factor), as well as to the different universe of test assets (portfolios test assets versus individual equities).

[Table 7 about here.]

5 A narrative climate-related U.S policy news factor

We have explained the finding that the U.S. climate policy textual factor carries a positive risk premium by using a hedging climate risk argument. For this to hold, an increase in the textual U.S. climate policy factor should signify good news for the economy, i.e. a decrease in transition risks. We have conjectured that this interpretation is valid by informally arguing that after 2012, most news has signalled a fall in transition risks for the U.S. economy. In this Section, we check whether the hedging argument explanation holds by creating a U.S. climate policy factor whose increase (decrease) signifies an increase (decrease) in transition risks by construction. Then, according to our hedging explanation, the factor should command a negative risk premium.

We construct our factor by taking a narrative approach which accounts for the content of each news article related to the U.S. climate topic. First, we select articles with a loading on the domestic policy factor greater than 40%; this yields 3,500 articles. We read each one of the 3,500 articles covering the topic of U.S. policy news and mark it with a 1 if it signals an increase in transition risks, with a -1 if it signals a fall, and with a zero if its content is mixed or only marginally related to the topic. Then, we create a time series capturing the transition risks elicited by the U.S. political debate by summing the marks given to the articles over each day.

Figure 3 shows the time series of climate change news based on the narrative approach. It reports monthly averages of the markings assigned at a daily frequency. Note that values close to zero do not necessarily imply that there were no news in a given month. Rather, they could indicate that daily news signalling an increase and a decrease of transition risk cancel out on average over a month. We identify four main periods based on the patterns of our time series. The first period spans January 2000-November 2006. Over this period, our narrative variable hovers around zero, revealing either a lack of interest from the government administration in tackling issues related to climate change, and/or a mix of positive and negative news for the economy which were cancelling out. This period corresponds to the administration of George W. Bush, until the Republicans lost the majority in the House of Representatives in November 2006. Over this period, the Republican party controlled both the House and the Senate, so President Bush was free to lead his political agenda on climate change.

[Figure 3 about here.]

The second period spans November 2006-November 2010, over which our narrative variable often takes positive values, signalling higher transition risks. This is a period where the Democratic party controls the House of Representative, and it is characterized by the administration of George W. Bush until November 2008 and that of Barak Obama afterwards. The third period spans November 2012 to November 2016, over which the time series of transition risks hovers again around zero, in a way that closely resembles the period of Bush' administration.This period is instead characterized by Obama's loss of control over the Congress. In November 2012, the Democratic party lost the majority in the House of Representatives, and in November 2014 it also lost the majority in the Senate. Over this period of time, the news reveals the inability of President Obama to tackle climate change, with his efforts of passing executive orders being offset by the strategies of the Republicans in the Congress. This is reflected in the observed pattern of our variable. Finally, the fourth period starting in November 2016 covers the Trump's administration, which was clearly characterized by a very pronounced fall in transition risks. Overall, the pattern of the time series in Figure 3 verifies our conjecture that after November 2012, the news coverage of U.S. climate policy tends to reflect a fall in transition risks, which becomes most pronounced after November 2016.

Next, we explore whether the U.S. climate policy narrative factor is priced. Given that an increase in the factor signifies an increase in transition risks by construction, it should command a negative risk premium, should our hedging perspective explanation holds. Table 8 reports the alphas of spread portfolios constructed from portfolio sorts with respect to the narrative measure of climate risks. We report results for decile and quintile portfolios across model specifications, over the full period and over 2000-2012 and 2012-2018 sub-samples. We can see that results are consistent with these reported in Section 4 in Tables 2 and 3. The narrative factor is priced in the post-2012 period in all cases, whereas it is not priced in the pre-2012 period. This explains why the evidence of it being priced over 2000-2018 is mixed. Moreover, alphas are negative. The results confirm the hedging argument as an explanation for the reported positive (negative) risk premium of the textual (narrative) U.S. climate policy factor. Stocks which are more exposed to the textual (narrative) factor are riskier because a decrease (increase) in the factor signifies an increase in transition risks. To hedge the risk of the textual (narrative) factor, investors buy stocks with negative (positive) climate betas, thus increasing their prices and lowering their returns. As a result, the long-short spread portfolio formed with respect to the textual (narrative) factor will yield a positive (negative) alpha, just as we find. The analysis based on the narrative approach corroborates the conclusion that the transition risks elicited by the U.S. political debate on climate change have only started to be priced over the most recent years.

[Table 8 about here.]

6 Conclusions

We examine whether climate change risks are priced in the cross-section of U.S. stocks over 2000-2018. We dissect climate change risk in its multiple sources, including physical and transition, short- and long-run risks, by using the LDA textual analysis method. LDA is a natural choice for the purposes of our analysis. It decomposes climate change risk by grouping words in categories (topics) which can have a clear interpretation. Furthermore, LDA delivers the association of any article with any given topic (topic share). The time series of topic share constitutes a risk factor associated with any given topic.

We identify four topics which have the clear interpretation of natural disasters, global warming, international summits, and U.S. climate policy, respectively. We find that only the risk related to U.S. climate policy is priced, with investors requiring a greater expected return for stocks being more exposed to this risk factor. This evidence is pronounced post-2012. We attribute our findings to the fact that investors hedge U.S. climate policy risks, in accordance with the assumptions of Pastor et al. (2020). We validate our explanation by constructing a U.S. climate policy narrative factor which confirms the asset pricing results obtained from the corresponding textual factor.

Our findings have at least two implications for investors and policymakers. First, climate change risks have started to be priced only very recently in the U.S. stock market; this echoes the survey results in Krueger et al. (2020), in which a large percentage of institutional investors also mention that climate change risks have come to their attention only in the last decade. Second, only a limited set of risks are priced. Climate change risks appear to be priced only when they reach the domestic U.S. political debate on climate. This implies that only the U.S. political arena serves as a “wake up” call to investors on climate change risk, and it suggests that longer term climate change risks may be substantially understated in U.S. stock market valuations. This calls for government intervention, especially given that these risks may materialize by 2050 (Intergovernmental Panel on Climate Change (2018). Future research should explore further why investors do not take these risks into account in their valuations. One explanation could be that investors have short term horizons, which may be the case for some types of institutional investors, but not for others (e.g., pension funds). Another explanation could be that there is not much information transmitted to investors on these types of risk. In this case, policy makers should improve the transmitting information channels to correct this inefficiency.

References

Asaf, B., G. Matthew T., and L. Ryan (2019): "Disaster on the Horizon: The Price Effect of Sea Level Rise," *Journal of Financial Economics*, 134, 253 – 272.

Baldauf, M., L. Garlappi, and C. Yannelis (2020): "Does Climate Change Affect Real Estate Prices? Only If You Believe In It," *Review of Financial Studies*, 33, 1256–1295.

Bansal, R., D. Kiku, and M. Ochoa (2017): "Price of Long-Run Temperature Shifts in Capital Markets," Working paper, Duke University.

Berg, F., J. Kölbel, and R. Rigobon (2020): " Aggregate Confusion: The Divergence of ESG Ratings," Working paper, MIT Sloan.

Bessembinder, H. (2017): "Fossil Fuel Divestment and Its Potential Impacts On Students, Faculty and Other University and Pension Stakeholders," Working paper, Arizona State University.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Bolton, P. and M. Kacperczyk (2020a): "Carbon Premium around the World," Working paper, Imperial College, London.

——— (2020b): "Do Investors Care about Carbon Risk?" Working paper, NBER.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

CFA Institute (2020): "Climate Change Analysis in the Investment Process," Tech. rep.

Chang, B. Y., P. Christoffersen, and K. Jacobs (2013): "Market Skewness Risk and the Cross Section of Stock Returns," *Journal of Financial Economics*, 107, 46 – 68.

Choi, D., Z. Gao, and W. Jiang (2020): "Attention to Global Warming," *Review of Financial Studies*, 33, 1112–1145.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, J. Stroebel, and A. Karolyi (2020): "Hedging Climate Change News," *Review of Financial Studies*, 33, 1184–1216.

Fama, E. F. and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

——— (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116, 1 – 22.

Fama, E. F. and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Gibson, R., P. Krueger, and S. Mitali (2020): "The Sustainability Footprint of Institutional Investors: ESG Driven Price Pressure and Performance," Working Paper 17-05, Swiss Finance Institute.

Griffiths, T. L. and M. Steyvers (2004): "Finding Scientific Topics," *Proceedings of the National Academy of Sciences*, 101, 5228–5235.

Görgen, M., A. Jacob, M. Nerlinger, R. Riordan, M. Rohleder, and M. Wilkens (2019): "Carbon Risk," Working paper, University of Augsburg.

Hansen, S., M. McMahon, and A. Prat (2017): "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach," *Quarterly Journal of Economics*, 133, 801–870.

Harvey, C. R. (2017): "Presidential Address: The Scientific Outlook in Financial Economics: Scientific Outlook in Finance," *Journal of Finance*, 72, 1399–1440.

Heinrich, G. (2009): "Parameter Estimation for Text Analysis," Tech. rep., Fraunhofer IGD.

Hong, H., F. W. Li, and J. Xu (2019): "Climate Risks and Market Efficiency," *Journal of Econometrics*, 208, 265 – 281, special Issue on Financial Engineering and Risk Management.

Hou, K., C. Xue, and L. Zhang (2020): "Replicating Anomalies," *Review of Financial Studies*, 33, 2019–2133.

Hsu, P.-H., K. Li, and C.-Y. Tsou (2020): "The Pollution Premium," Working paper.

Ilhan, E., Z. Sautner, and G. Vilkov (2020): "Carbon Tail Risk," *Review of Financial Studies*, forthcoming.

In, S. Y., K. Y. Park, and A. H. B. Monk (2019): "Is 'being Green' Rewarded in the Market?: An Empirical Investigation of Decarbonization and Stock Returns," Working paper, Stanford Global Project Center.

Intergovernmental Panel on Climate Change (2018): "Summary for Policymakers." in *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, 85–113, [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.

International Monetary Fund (2020): "Mitigating Climate Change—Growth- and Distribution-Friendly Strategies," in *World Economic Outlook*, 85–113, chapter 3.

Krueger, P., Z. Sautner, and L. T. Starks (2020): "The Importance of Climate Risks for Institutional Investors," *Review of Financial Studies*, 33, 1067–1111.

Li, Q., H. Shan, Y. Tang, and V. Yao (2020): "Corporate Climate Risk: Measurements and Responses ," Working paper, European Corporate Governance Institute.

LOUGHRAN, T. and B. MCDONALD (2011): "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, 66, 35–65.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Painter, M. (2020): "An Inconvenient Cost: The Effects of Climate Change on Municipal Bonds," *Journal of Financial Economics*, 135, 468 – 482.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2020): "Sustainable Investing in Equilibrium," *Journal of Financial Economics*, forthcoming.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski (2020): "Responsible Investing: The ESG-Efficient frontier," *Journal of Financial Economics*, forthcoming.

Romer, C. D. and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang (2020): "Firm-level Climate Change Exposure," Working paper, European Corporate Governance Institute.

Shumway, T. (1997): "The Delisting Bias in CRSP Data," *The Journal of Finance*, 52, 327–340.

Thorsrud, L. A. (2016): "Nowcasting Using News Topics. Big Data versus Big Bank," Working Paper 20, Norges Bank.

Zerbib, O. D. (2020): "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Green Investing and Sin Stock Exclusion," Working paper, Tilburg University.

Zhao, W., J. J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, and W. Zou (2015): "A Heuristic Approach to Determine an Appropriate Number of Topics in Topic Modeling," *BMC Bioinformatics*, 16, S8.

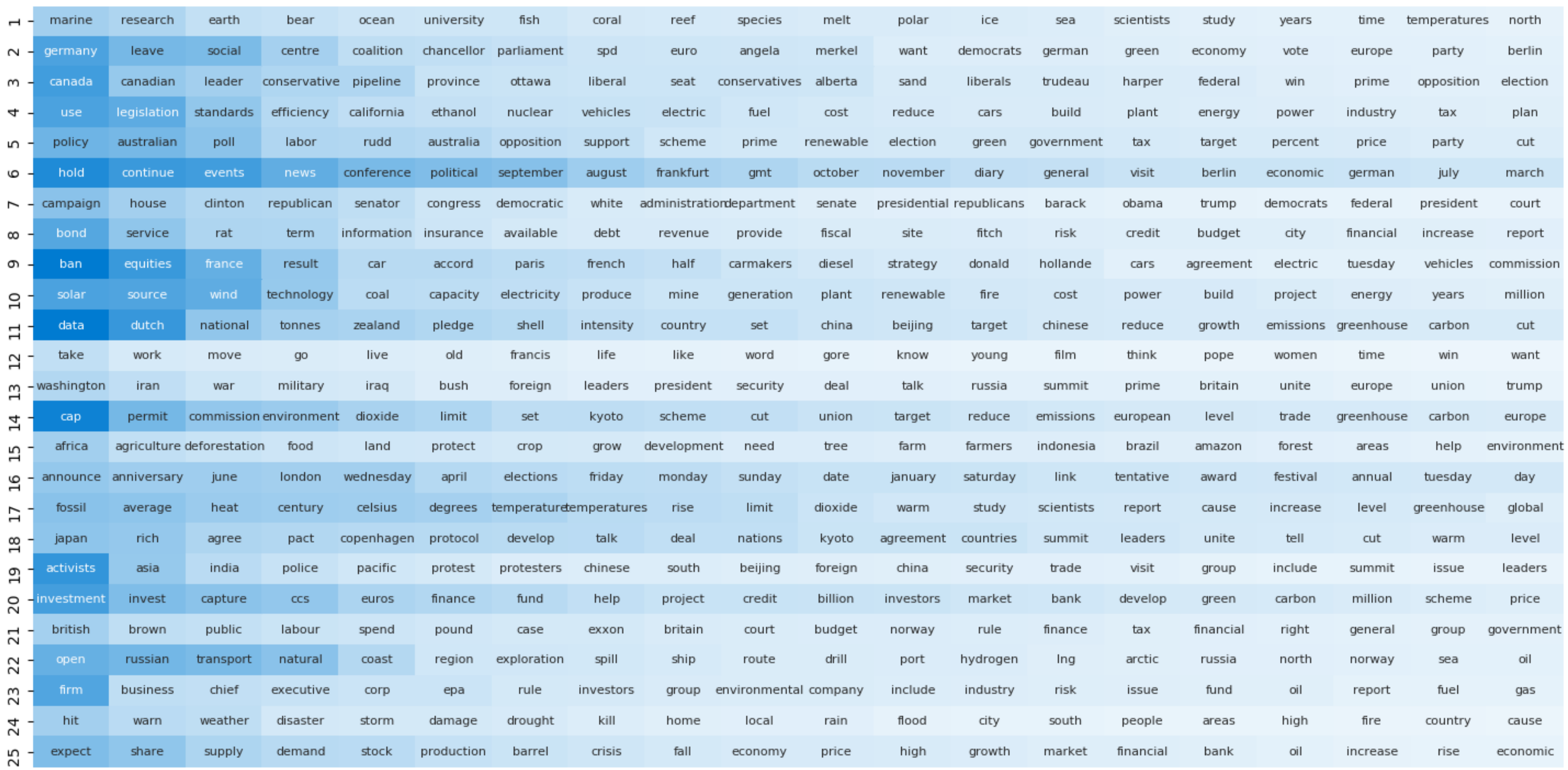
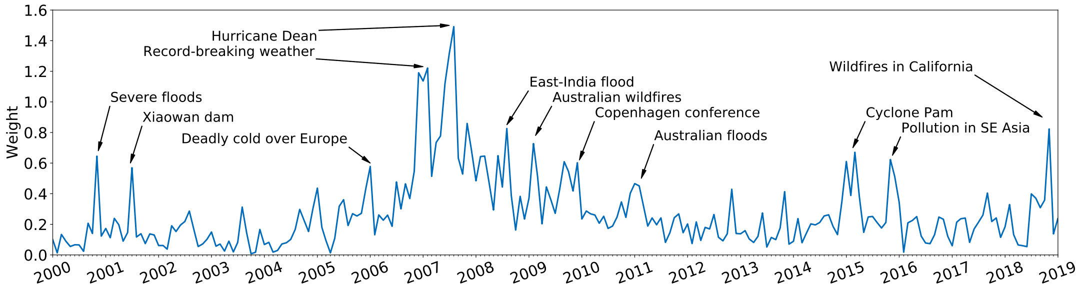
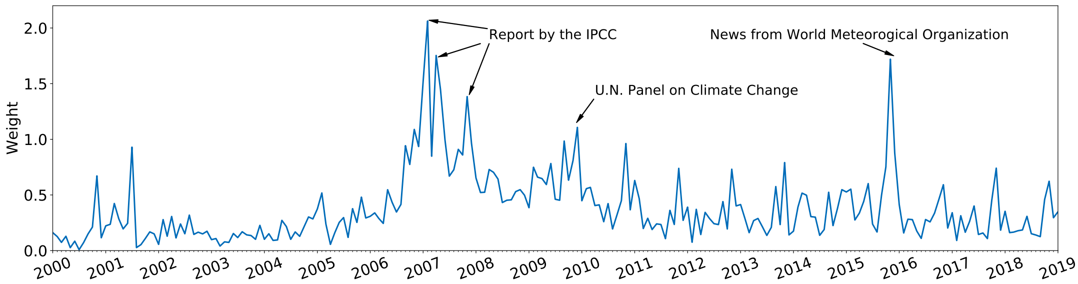


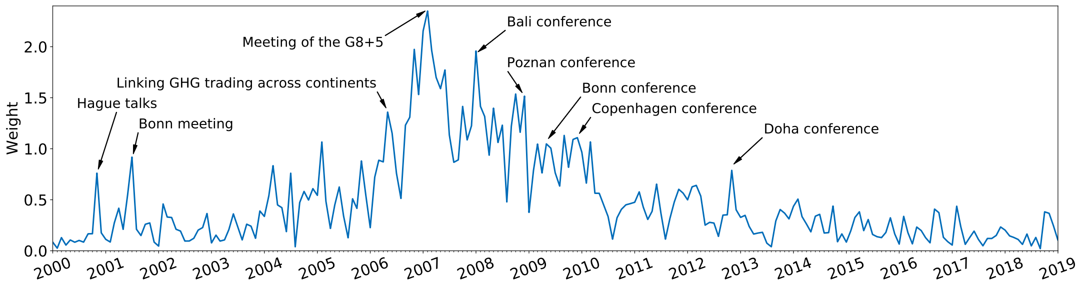
Figure 1. Topics and Terms Within Topics Ranked by Probability.



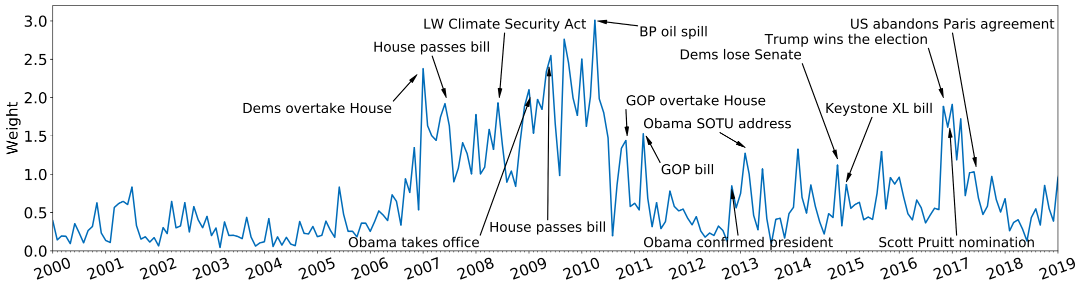
(a) Natural disasters textual factor.



(b) Global warming textual factor.



(c) International summits textual factor



(d) U.S. climate policy textual factor.

Figure 2. Climate textual factors over January 2000 - December 2018 and their association with news releases.

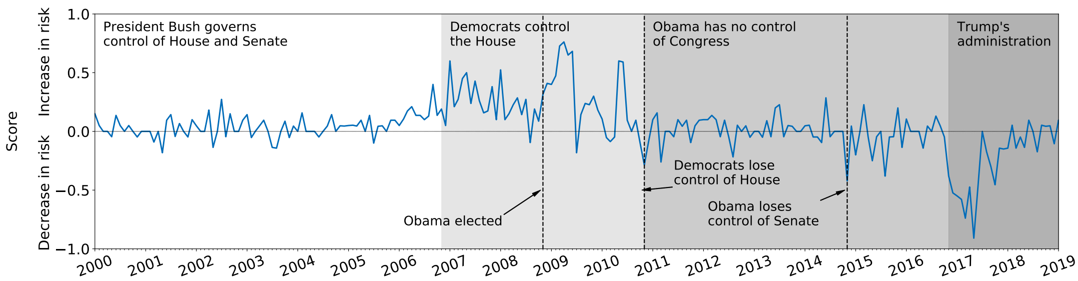


Figure 3. A Narrative Measure of U.S. Climate Policy Risks. The figure reports the monthly averages of the markings assigned at a daily frequency to each one of 3,500 articles related to U.S. climate policy. An increase (decrease) in the factor signifies an increase (decrease) in transition risks.

Table 1. Pairwise correlations between textual factors and other risk factors, January 1, 2000 - December 31, 2018.

	U.S. climate policy	Int'l summits	Global warming	Natural disasters	mktrf	hml	smb	rmw	cma	umd
U.S. domestic policy	1.00	0.30	0.27	0.18	−0.02	−0.02	0.01	0.02	−0.02	−0.00
International summits	0.30	1.00	0.31	0.24	−0.01	0.01	0.00	0.02	−0.01	−0.00
Global warming	0.27	0.31	1.00	0.34	−0.01	−0.01	−0.01	0.02	−0.01	0.01
Natural disasters	0.18	0.24	0.34	1.00	−0.02	−0.03	−0.02	0.02	−0.01	0.04
mktrf	−0.02	−0.01	−0.01	−0.02	1.00	0.06	0.12	−0.44	−0.28	−0.30
hml	0.02	0.01	−0.01	−0.03	0.06	1.00	−0.18	0.06	0.45	−0.33
smb	0.01	0.00	−0.01	−0.02	0.12	−0.18	1.00	−0.35	−0.05	0.13
rmw	0.02	0.02	0.02	0.02	−0.44	0.06	−0.35	1.00	0.26	0.17
cma	−0.02	−0.01	−0.01	−0.01	−0.28	0.45	−0.05	0.26	1.00	0.11
umd	−0.00	−0.00	0.01	0.04	−0.30	−0.33	0.13	0.17	0.11	1.00

Table 2. Portfolio sort analysis: Climate textual factors, January 1, 2000 - December 31, 2018

	U.S. Climate	International Summits	Global Warming	Natural Disasters
Panel A: Market model				
Deciles	0.96***	−0.12	−0.08	0.14
	(2.91)	(−0.42)	(−0.28)	(0.38)
Quintiles	0.59**	−0.17	0.31	0.06
	(2.31)	(−0.70)	(1.46)	(0.17)
Panel B: Fama-French three factor model				
Deciles	0.65**	−0.53*	0.20	0.07
	(2.34)	(−1.73)	(0.67)	(0.24)
Quintiles	0.30	−0.25	0.09	0.01
	(1.24)	(−1.21)	(0.55)	(0.04)
Panel C: Fama-French-Carhart model				
Deciles	0.46*	−0.49	0.03	−0.07
	(1.66)	(−1.65)	(0.10)	(−0.24)
Quintiles	0.09	−0.14	0.27*	0.06
	(0.46)	(−0.71)	(1.92)	(0.38)
Panel D: Fama-French five factor model				
Deciles	0.82***	−0.66**	0.05	0.03
	(2.75)	(−2.58)	(0.19)	(0.08)
Quintiles	0.54***	−0.18	0.13	0.04
	(2.63)	(−0.96)	(0.67)	(0.19)
Panel E: Fama-French five factor model plus momentum factor				
Deciles	0.61**	−0.76***	−0.09	0.27
	(2.25)	(−2.63)	(−0.34)	(0.89)
Quintiles	0.30**	−0.16	0.22	0.10
	(1.99)	(−0.86)	(1.20)	(0.53)

Notes: Entries report the alpha of the spread portfolio, estimated from monthly post-ranking returns, over January 1, 2000 - December 31, 2018; the unit is % per month. At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to a given climate textual factor (global warming, natural disasters, international summits and U.S. climate policy textual factors). Then, we compute the post-ranking value-weighted portfolio monthly return over the period t to $t+1$. The resulting spread's portfolio return at $t+1$ is computed as the difference between the return of portfolio 10 (high climate beta) minus the return of portfolio 1 (low climate beta). A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. Betas of stocks and alpha of the spread portfolio are estimated by the same set of control variables X_t in equation 1. We use five alternative specifications. The market model includes only the market factor. FF3 denotes the Fama-French (Fama and French (1993)) three factor model, which includes the market, size and book to market factors. FFC is the four factor Fama-French-Carhart (Carhart (1997)) model, that adds a momentum factor to the controls in FF3. FF5 is the Fama-French five factor model (Fama and French (2015)), that includes investment and profitability factors in addition to the controls in FF3. FF5+ umd is a model that includes the momentum factor in addition to the controls in FF5. Newey and West (1987) t-statistics with 6 lags are reported in parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

Table 3. Portfolio sort analysis: Climate policy Factor, Pre- and Post-Obama's second election

	U.S. Climate		International Summits		Global Warming		Natural Disasters	
	Pre-2012	Post-2012	Pre-2012	Post-2012	Pre-2012	Post-2012	Pre-2012	Post-2012
Panel A: Market model								
Deciles	1.05**	0.84**	−0.17	−0.14	−0.03	−0.04	0.19	0.22
	(2.33)	(2.12)	(−0.49)	(−0.32)	(−0.09)	(−0.09)	(0.42)	(0.39)
Quintiles	0.55	0.75***	−0.27	0.11	0.47*	−0.05	0.23	−0.14
	(1.55)	(2.89)	(−0.89)	(0.32)	(1.77)	(−0.14)	(0.53)	(−0.55)
Panel B: Fama-French three factor model								
Deciles	0.35	0.98***	−0.80**	−0.37	0.18	0.33	−0.05	0.70
	(0.91)	(3.06)	(−2.05)	(−0.78)	(0.46)	(0.73)	(−0.15)	(1.30)
Quintiles	0.06	0.70***	−0.52*	0.12	−0.03	0.16	0.01	0.21
	(0.17)	(3.11)	(−1.94)	(0.38)	(−0.13)	(0.60)	(0.04)	(0.64)
Panel C: Fama-French-Carhart model								
Deciles	0.17	0.97***	−0.66*	−0.77	−0.13	0.30	−0.15	0.58
	(0.46)	(3.29)	(−1.83)	(−1.59)	(−0.30)	(0.75)	(−0.54)	(1.05)
Quintiles	−0.11	0.46**	−0.31	−0.10	0.10	0.28	0.13	−0.02
	(−0.43)	(2.52)	(−1.23)	(−0.40)	(0.59)	(1.23)	(0.69)	(−0.06)
Panel D: Fama-French five factor model								
Deciles	0.48	1.23***	−1.07***	−0.19	0.04	0.38	0.47	−0.09
	(1.23)	(3.82)	(−3.37)	(−0.48)	(0.12)	(0.86)	(0.91)	(−0.27)
Quintiles	0.45*	0.59**	−0.48**	0.31	0.24	0.05	0.10	0.03
	(1.73)	(2.15)	(−2.05)	(1.09)	(1.00)	(0.20)	(0.37)	(0.10)
Panel E: Fama-French five factor model plus momentum factor								
Deciles	0.44	0.79***	−1.23***	−0.55	0.27	−0.20	0.50	0.24
	(1.26)	(2.72)	(−3.42)	(−1.28)	(0.75)	(−0.59)	(1.08)	(0.84)
Quintiles	0.21	0.42**	−0.39	−0.09	0.14	0.25	0.15	0.18
	(1.12)	(2.13)	(−1.56)	(−0.38)	(0.65)	(0.98)	(0.67)	(0.76)

Notes: Entries report the alpha of the spread portfolio, estimated from monthly post-ranking returns, over the sub-periods January 1, 2000 - November 5, 2012 and November 6, 2012 - December 31, 2018; the unit is % per month. At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to a given climate textual factor (global warming, natural disasters, international summits and U.S. climate policy textual factors). Then, we compute the post-ranking value-weighted portfolio monthly return over the period t to $t+1$. The resulting spread portfolio return at $t+1$ is computed as the difference between the return of portfolio 10 (high climate beta) minus the return of portfolio 1 (low climate beta). A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. Betas of stocks and alpha of the spread portfolio are estimated by the same set of control variables X_t in equation 1. We use five alternative specifications. The market model includes only the market factor. FF3 denotes the Fama-French (Fama and French (1993)) three factor model, which includes the market, size and book to market factors. FFC is the four factor Fama-French-Carhart (Carhart (1997)) model, that adds a momentum factor to the controls in FF3. FF5 is the Fama-French five factor model (Fama and French (2015)), that includes investment and profitability factors in addition to the controls in FF3. FF5+ umd is a model that includes the momentum factor in addition to the controls in FF5. Newey and West (1987) t-statistics with 6 lags are reported within parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

Table 4. Portfolio characteristics

	1 (L)	2	3	4	5 (H)
Panel A: Market model					
Return	0.71	0.86**	0.90***	0.98***	1.26***
	(1.52)	(2.34)	(2.95)	(2.89)	(3.19)
Climate β	−0.49	−0.17	−0.01	0.16	0.47
ESG	36.42	41.37	42.14	41.03	36.12
ESG (change)	8.31	6.02	6.30	6.50	6.50
ESG coverage	26.74	44.33	47.56	45.11	28.26
log(size)	6.43	6.91	6.98	6.92	6.48
N	746	749	752	750	746
Panel B: Fama-French three factor model					
Return	0.60	0.97***	0.87***	0.97***	1.23***
	(1.27)	(2.72)	(2.85)	(3.05)	(2.92)
Climate β	−0.48	−0.16	0.00	0.16	0.48
ESG	36.32	41.51	42.36	41.11	35.84
ESG (change)	7.36	6.56	5.80	6.56	6.48
ESG coverage	26.45	45.24	48.13	44.93	27.26
log(size)	6.39	6.91	7.01	6.91	6.43
N	747	750	751	751	746
Panel C: Fama-French-Carhart model					
Return	0.80*	1.03***	0.87***	0.89**	1.07***
	(1.84)	(2.88)	(2.84)	(2.60)	(2.66)
Climate β	−0.48	−0.16	0.00	0.15	0.47
ESG	36.05	41.20	42.56	41.23	35.82
ESG (change)	7.81	6.51	5.90	6.55	6.16
ESG coverage	25.87	45.29	48.44	45.29	27.12
log(size)	6.36	6.91	7.02	6.91	6.43
N	747	751	751	750	747
Panel D: Fama-French five factor model					
Return	0.71	1.01***	0.86***	0.95***	1.10***
	(1.40)	(2.76)	(2.79)	(3.09)	(2.93)
Climate β	−0.48	−0.16	0.00	0.16	0.48
ESG	36.08	41.43	42.18	41.27	36.23
ESG (change)	7.19	6.48	5.84	6.68	6.39
ESG coverage	26.17	45.15	48.24	45.16	27.27
log(size)	6.38	6.92	7.01	6.91	6.43
N	747	748	752	752	747
Panel E: Fama-French five factor model plus momentum factor					
Return	0.83*	1.04***	0.88***	0.85***	1.04***
	(1.81)	(2.78)	(2.85)	(2.70)	(2.90)
Climate β	−0.48	−0.16	0.00	0.16	0.47
ESG	35.89	41.20	42.49	41.20	36.16
ESG (change)	7.31	6.54	5.91	6.79	5.91
ESG coverage	25.84	45.09	48.72	45.48	26.87
log(size)	6.36	6.91	7.01	6.91	6.42
N	746	750	753	751	747

Notes: Entries report the average portfolio climate beta, average value-weighted portfolio return, average environmental pillar indicator from Thomson Reuters ESG scores, average percentage yearly change in the environment pillar indicator, the average market capitalisation (log size), and the number N of firms included in each decile portfolio. One, two, and three asterisks indicate significance at a 10%, 5% and 1% level, respectively. All statistics refer to the period November 2012-December 2018.

Table 5. Portfolio sort analysis: Aggregate climate textual factor, January 1, 2000 - December 31, 2018

	Aggregate factor
Panel A: Market model	
Deciles	0.21
	(0.68)
Quintiles	0.35
	(1.34)
Panel B: Fama-French three factor model	
Deciles	0.11
	(0.43)
Quintiles	0.17
	(0.84)
Panel C: Fama-French-Carhart model	
Deciles	0.07
	(0.26)
Quintiles	0.03
	(0.19)
Panel D: Fama-French five factor model	
Deciles	0.23
	(0.71)
Quintiles	0.46**
	(2.43)
Panel E: Fama-French five factor model plus momentum factor	
Deciles	0.07
	(0.26)
Quintiles	0.30*
	(1.89)

Notes: Entries report the alpha of the spread portfolio, estimated from monthly post-ranking returns, over January 1, 2000 - December 31, 2018; the unit is % per month. At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to a given climate textual factor (global warming, natural disasters, international summits and U.S. climate policy textual factors). Then, we compute the post-ranking value-weighted portfolio monthly return over the period t to $t+1$. The resulting spread portfolio return at $t+1$ is computed as the difference between the return of portfolio 10 (high climate beta) minus the return of portfolio 1 (low climate beta). A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. Betas of stocks and alpha of the spread portfolio are estimated by the same set of control variables X_t in equation 1. We use five alternative specifications. The baseline model includes only the market factor. FF3 denotes the Fama-French (Fama and French (1993)) three factor model, which includes the market, size and book to market factors. FFC is the four factor Fama-French-Carhart (Carhart (1997)) model, that adds a momentum factor to the controls in FF3. FF5 is the Fama-French five factor model (Fama and French (2015)), that includes investment and profitability factors in addition to the controls in FF3. FF5+ umd is a model that includes the momentum factor in addition to the controls in FF5. Newey and West (1987) t-stats with 6 lags are reported in parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

Table 6. Fama and MacBeth (1973) regressions

	55FF					74FF				
	(i)	(ii)	(iii)	(iv)	(v)	(i)	(ii)	(iii)	(iv)	(v)
Panel A: 2000-2012										
mktrf	−0.01***	−0.01***	−0.01***	−0.01***	−0.01***	−0.01**	−0.01**	−0.01***	−0.01***	−0.01***
	(−2.67)	(−2.78)	(−2.90)	(−2.67)	(−3.02)	(−2.17)	(−2.40)	(−2.63)	(−2.37)	(−2.86)
hml	0.00	0.00**	0.00	0.00**	0.00*	0.00*	0.00	0.00**	0.00*	0.00*
	(1.15)	(1.19)	(1.26)	(0.90)	(0.97)	(1.62)	(1.74)	(1.68)	(1.52)	(1.44)
smb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.52)	(0.74)	(0.40)	(0.54)	(0.65)	(0.33)	(0.49)	(0.23)	(0.28)	(0.45)
umd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(−0.65)	(−0.38)	(−0.37)	(−0.69)	(−0.37)	(−0.85)	(−0.75)	(−0.72)	(−0.92)	(−0.52)
Natural Disasters	−0.05				−0.07	0.11				−0.02
	(−0.13)				(−0.18)	(0.42)				(−0.06)
Global Warming		0.26			0.32		0.40			0.30
		(0.47)			(0.68)		(0.98)			(0.86)
International Summits			0.16		−0.16			−0.01		−0.30
			(0.23)		(−0.24)			(−0.01)		(−0.39)
U.S. Climate Policy				1.02*	0.29				0.57	0.02***
				(1.37)	(0.38)				(0.83)	(0.03)
Panel B: 2012-2018										
mktrf	−0.01**	−0.01***	−0.01***	−0.01**	−0.01***	−0.01**	−0.01**	−0.01**	−0.01*	−0.01**
	(−2.37)	(−2.75)	(−2.23)	(−1.83)	(−2.41)	(−1.96)	(−1.91)	(−1.48)	(−1.48)	(−1.86)
hml	0.00*	0.00**	0.00*	0.00**	0.00*	0.00	0.00	0.00*	0.00*	0.00
	(−1.56)	(−1.63)	(−1.33)	(−1.92)	(−1.49)	(−1.19)	(−1.43)	(−1.22)	(−1.59)	(−1.31)
smb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(−0.42)	(−0.54)	(−0.52)	(−0.30)	(0.13)	(−0.53)	(−0.64)	(−0.83)	(−0.64)	(−0.07)
umd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.44)	(−0.07)	(0.1)	(0.26)	(0.45)	(0.76)	(0.60)	(0.53)	(0.55)
Natural Disasters	−0.04				−0.72**	−0.09				−0.65
	(−0.08)				(−1.91)	(−0.18)				(−1.13)
Global Warming		−0.22			0.04		−0.18			−0.03
		(0.35)			(0.08)		(−0.35)			(−0.56)
International Summits			0.61***		0.20			0.01		−0.08
			(2.30)		(0.65)			(0.04)		(−0.30)
U.S. Climate Policy				2.28***	2.11***				2.12***	1.84***
				(2.52)	(2.45)				(2.55)	(2.36)

Notes: Entries report the prices of risks obtained from Fama and MacBeth (1973) regressions (FM) over the 2012-2018 period. We apply FM regressions to the 55 and 74 Fama-French industry portfolios, separately.In the first-pass regressions, for each security, we estimate climate betas using a rolling window of the daily observations over the past three months. We repeat the procedure by rolling the beta estimation window by one month, just as we did in the portfolio-sort approach to asset pricing tests. We estimate factor betas by Carhart (1997) model. In the second pass regressions, at each time step, we obtain the price of risk of each factor by running cross-sectional regressions of the stock returns over the next month on the estimated betas of the factors obtained from the first-pass regressions. Specifications (i), (ii), (iii), (iv) and (v), consider the four textual factors (natural disasters, global warming, international summits, U.S. climate policy) separately and jointly, respectively, while controlling for the Carhart (1997) factors. Newey and West (1987) t-statistics with 6 lags are reported within parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

Table 7. Portfolio sort analysis over annual horizons: Climate textual factors, January 1, 2000 - December 31, 2018

	U.S. Climate	International Summits	Global Warming	Natural Disasters
Panel A: Market model				
Deciles	2.96	−0.47	1.48	1.91
	(1.24)	(−0.28)	(0.96)	(1.03)
Quintiles	0.71	0.80	2.94**	1.92
	(0.47)	(0.54)	(2.51)	(1.45)
Panel B: Fama-French three factor model				
Deciles	1.67	−0.94	1.65	0.37
	(0.87)	(−0.50)	(1.54)	(0.25)
Quintiles	1.11	−0.93	1.33	0.31
	(0.85)	(−0.72)	(1.04)	(0.28)
Panel C: Fama-French-Carhart model				
Deciles	1.30	0.47	1.11	−0.85
	(0.75)	(0.25)	(1.06)	(−0.56)
Quintiles	0.62	0.24	1.74	0.17
	(0.55)	(0.19)	(1.60)	(0.16)
Panel D: Fama-French five factor model				
Deciles	2.41	0.14	0.91	−0.35
	(1.46)	(0.08)	(0.76)	(−0.20)
Quintiles	1.43	−0.55	1.22	−0.09
	(1.28)	(−0.55)	(1.09)	(−0.09)
Panel E: Fama-French five factor model plus momentum factor				
Deciles	2.31	0.11	2.18**	−1.29
	(1.34)	(0.06)	(2.08)	(−0.82)
Quintiles	0.14	−0.03	1.52*	−0.46
	(0.14)	(−0.03)	(1.77)	(−0.42)

Notes: Entries report the alpha of the spread portfolio, estimated from annual overlapping post-ranking returns, over January 1, 2000 - December 31, 2018; the unit is % per month. At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to a given climate textual factor (global warming, natural disasters, international summits and U.S. climate policy textual factors). Then, we compute the post-ranking value-weighted portfolio annual return. The resulting spread portfolio return is computed as the difference between the return of portfolio 10 (high climate beta) minus the return of portfolio 1 (low climate beta). A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. The betas of stocks and alpha of the spread portfolio are estimated by the same set of control variables X_t in equation 1. We use five alternative specifications. The baseline model includes only the market factor. FF3 denotes the Fama-French (Fama and French (1993)) three factor model, which includes the market, size and book to market factors. FFC is the four factor Fama-French-Carhart (Carhart (1997)) model, that adds a momentum factor to the controls in FF3. FF5 is the Fama-French five factor model (Fama and French (2015)), that includes investment and profitability factors in addition to the controls in FF3. FF5+ umd is a model that includes the momentum factor in addition to the controls in FF5. Newey and West (1987) t-statistics with 6 lags are reported in parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

Table 8. Narrative factor: Portfolio sort analysis over subsamples

	2000-2018	2000-2012	2012-2018
Panel A: Market model			
Deciles	-0.64^{*}	-0.52	-1.01^{**}
	(-1.86)	(-1.13)	(-2.43)
Quintiles	-0.23	0.01	-0.71
	(-0.77)	(0.02)	(-1.52)
Panel B: Fama-French three factor model			
Deciles	-1.03^{***}	-0.77^{**}	-1.39^{***}
	(-3.56)	(-2.37)	(-4.30)
Quintiles	-0.58^{***}	-0.20	-1.05^{***}
	(-2.64)	(-0.78)	(-3.67)
Panel C: Fama-French-Carhart model			
Deciles	-0.85^{***}	-0.59^{*}	-1.37^{***}
	(-2.76)	(-1.66)	(-3.61)
Quintiles	-0.48^{**}	-0.24	-0.93^{***}
	(-2.30)	(-1.05)	(-2.86)
Panel D: Fama-French five factor model			
Deciles	-0.65^{**}	-0.62	-0.84^{***}
	(-1.97)	(-1.43)	(-2.97)
Quintiles	-0.39^{*}	-0.16	-0.69^{**}
	(-1.89)	(-0.62)	(-2.53)
Panel E: Fama-French five factor model plus momentum			
Deciles	-0.31	0.00	-0.93^{***}
	(-1.07)	(0.00)	(-3.40)
Quintiles	-0.26	0.05	-0.60^{**}
	(-1.20)	(0.19)	(-2.08)

Notes: Entries report the alpha of the spread portfolio, estimated from monthly post-ranking returns, over January 2000 - December 2018, January 2000 - November 2012, and November 2012 - December 2018; the unit is % per month. At the end of each month t, we sort stocks in ascending order in decile portfolios, based on the magnitude of their estimated climate betas with respect to the narrative U.S. climate policy factor. Then, we compute the post-ranking value-weighted portfolio monthly return. The resulting spread portfolio return is computed as the difference between the return of the portfolio with the highest climate beta minus the return of the portfolio with the lowest climate beta. A rolling window of daily observations over the past three months is used to estimate climate betas, and the window is rolled forward by one-month at each estimation step. The betas of stocks and alpha of the spread portfolio are estimated by the same set of control variables X_t in equation 1. We use five alternative specifications. The baseline model includes only the market factor. FF3 denotes the Fama-French (Fama and French (1993)) three factor model, which includes the market, size and book to market factors. FFC is the four factor Fama-French-Carhart (Carhart (1997)) model, that adds a momentum factor to the controls in FF3. FF5 is the Fama-French five factor model (Fama and French (2015)), that includes investment and profitability factors in addition to the controls in FF3. FF5+ umd is a model that includes the momentum factor in addition to the controls in FF5. Newey and West (1987) t-statistics with 6 lags are reported in parentheses. One, two, and three stars indicate 10%, 5% and 1% significance, respectively.

A Latent Dirichlet Allocation for Topic Identification

To process the news articles, we follow standard procedures. We first remove punctuation marks, newlines and tabs, and convert letters to lower case. Then we remove stop words (such as *the*, *is*, *are*, and *this*) and lemmatize all words, where the purpose of the latter is to reduce words to their respective word stems in order to limit the textual variability across documents. Finally, we trim the corpus such that tokens that occur less than 15 times and in more than 50% of the documents are removed in order to filter tokens that are either very rare or typical. This procedure returns a final dictionary with approximately 3 000 tokens.

Estimation Procedure Latent Dirichlet allocation (LDA) is conceptually a relatively simple procedure, yet computationally infeasible to estimate exactly due to the large discrete state space. Several approximate inference algorithms exist where the introductory LDA-paper by Blei et al. (2003) used a Variational Bayes approximation of the posterior distribution. A common alternative is collapsed Gibbs sampling which in the context of LDA was first employed by Griffiths and Steyvers (2004). We summarize the main idea behind this method by following the approach in Heinrich (2009) and Thorsrud (2016).

LDA defines a topic k to be a probability distribution $\boldsymbol{\phi}_k$ over all unique words in the textual corpus, i.e. the vocabulary V. An article m is represented as a probability distribution $\boldsymbol{\theta}_m$ over all possible topics. The probability distributions $\boldsymbol{\phi}_k$ and $\boldsymbol{\theta}_m$ are assumed to be drawn from Dirichlet distributions parameterized by the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, where bold letters denote vectors. LDA therefore assumes that the textual corpus is generated by the following process:

1. For each topic $k \in \{1, \ldots, K\}$
   (a) Draw a distribution over words $\boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\alpha})$
2. For each article $m \in \{1, \ldots, M\}$
   (a) Draw a vector of topic proportions $\boldsymbol{\theta}_m \sim \text{Dirichlet}(\boldsymbol{\beta})$

(b) For each word entry $n \in \{1, \ldots, N_m\}$

i. Draw a topic assignment $z_{m,n} \sim \text{Multinomial}(\boldsymbol{\theta}_m)$, $z_{m,n} \in \{1, \ldots, K\}$

ii. Draw a word $w_{m,n} \sim \text{Multinomial}(\boldsymbol{\phi}_{z_{m,n}})$

This generative procedure is illustrated graphically in Figure A.1. From this the following joint probability distribution of all known and hidden variables given hyperparameters for an article can be deduced

$$P(\boldsymbol{w}_m, \boldsymbol{z}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \overbrace{\underbrace{\prod_{n=1}^{N_m} P(w_{m,n}|\boldsymbol{\phi}_{z_{m,n}}) \times P(z_{m,n}|\boldsymbol{\theta}_m)}_{\text{word plate}} \times P(\boldsymbol{\theta}_m|\boldsymbol{\alpha})}^{\text{document plate}} \times \underbrace{P(\boldsymbol{\Phi}|\boldsymbol{\beta})}_{\text{topic plate}} \tag{A.1}$$

which across the entire textual sample of articles takes the form

$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{k=1}^{K} P(\boldsymbol{\phi}_k|\boldsymbol{\beta}) \times \prod_{m=1}^{M} P(\boldsymbol{\theta}_m|\boldsymbol{\alpha}) \times \prod_{n=1}^{N_m} P(z_{m,n}|\boldsymbol{\theta}_m) \times P(w_{m,n}|\boldsymbol{\phi}_{z_{m,n}}) \tag{A.2}$$

$$= P(\boldsymbol{\Phi}|\boldsymbol{\beta}) \times P(\boldsymbol{\Theta}|\boldsymbol{\alpha}) \times P(\boldsymbol{Z}|\boldsymbol{\Theta}) \times P(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) \tag{A.3}$$

We have used a capitalized bold letter to denote a matrix, e.g. $\boldsymbol{W} = \{\boldsymbol{w}_m\}_{m=1}^{M}$. The joint distribution can be found by integrating out the distributions $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ in (A.3), yielding

$$P(\boldsymbol{W}, \boldsymbol{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_{\boldsymbol{\Phi}} P(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) P(\boldsymbol{\Phi}|\boldsymbol{\beta})\, \mathrm{d}\boldsymbol{\Phi} \times \int_{\boldsymbol{\Theta}} P(\boldsymbol{Z}|\boldsymbol{\Theta}) P(\boldsymbol{\Theta}|\boldsymbol{\alpha})\, \mathrm{d}\boldsymbol{\Theta} \tag{A.4}$$

$$= P(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\beta}) \times P(\boldsymbol{Z}|\boldsymbol{\alpha}) \tag{A.5}$$

The two factors on the right-hand side of (A.5) are independent of $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, respectively. This allows for evaluating them separately, and using the properties of the Dirichlet distribution it can be shown that they take the form

$$P(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\beta}) = \prod_{k=1}^{K} \frac{\prod_{t=1}^{V}\Gamma(n_k^{(t)}+\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)} \frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\Gamma(\sum_{t=1}^{V} n_k^{(t)}+\beta_t)} \tag{A.6}$$

$$P(\boldsymbol{Z}|\boldsymbol{\alpha}) = \prod_{m=1}^{M} \frac{\prod_{k=1}^{K}\Gamma(n_m^{(k)}+\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\Gamma(\sum_{k=1}^{K} n_m^{(k)}+\alpha_k)} \tag{A.7}$$

where $n_k^{(t)}$ is the number of times the term t of the vocabulary has been assigned the kth topic and $n_m^{(k)}$ denotes the number of word tokens in the mth article assigned to the kth topic. These count structures have dimension $M \times K$ and $K \times V$, respectively.

[Figure 4 about here.]

The aim of the Gibbs sampler is to approximate the conditional distribution $P(\boldsymbol{Z}|\boldsymbol{W},\boldsymbol{\alpha},\boldsymbol{\beta})$ which is directly proportional to the joint distribution $P(\boldsymbol{W},\boldsymbol{Z}|\boldsymbol{\alpha},\boldsymbol{\beta})$ through the conditional probability

$$P(\boldsymbol{Z}|\boldsymbol{W},\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{P(\boldsymbol{W},\boldsymbol{Z}|\boldsymbol{\alpha},\boldsymbol{\beta})}{P(\boldsymbol{W}|\boldsymbol{\alpha},\boldsymbol{\beta})} \tag{A.8}$$

From the joint distribution $P(\boldsymbol{W},\boldsymbol{Z}|\boldsymbol{\alpha},\boldsymbol{\beta})$ we can derive the full conditional distribution for a word token $i \equiv (m,n)$, i.e. the update equation from which the Gibbs sampler draws the hidden variable z_i. It can be shown to take the form

$$P(z_i = k|\boldsymbol{Z}_{-i},\boldsymbol{W}) = \frac{P(\boldsymbol{W},\boldsymbol{Z})}{P(\boldsymbol{W},\boldsymbol{Z}_{-i})} \tag{A.9}$$

$$= \frac{P(\boldsymbol{W}|\boldsymbol{Z})}{P(\boldsymbol{W}_{-i}|\boldsymbol{Z}_{-i})P(w_i)} \times \frac{P(\boldsymbol{Z})}{P(\boldsymbol{Z}_{-i})} \tag{A.10}$$

$$\propto \frac{n_{k,-i}^{(t)}+\beta_t}{\sum_{t=1}^{V} n_{k,-i}^{(t)}+\beta_t} \times \frac{n_{m,-i}^{(k)}+\alpha_k}{\sum_{k=1}^{K} n_m^{(k)}+\alpha_k - 1} \tag{A.11}$$

where we have omitted the hyperparameters for clarity and the count $n_{\cdot,-i}^{(\cdot)}$ denotes that token i is excluded from the corresponding article or topic. Having initialized the topic assignments $z_{m,n}$ randomly from $\{1,\dots,K\}$, we can thus approximate the posterior distribution $P(\boldsymbol{Z}|\boldsymbol{W})$ by sampling the expression (A.11) for each word in an article and across all articles until a suitable steady state has been reached. Given a simulated sample of the posterior distribution, $\boldsymbol{\phi}_k$ and $\boldsymbol{\theta}_m$ can be estimated from

$$\phi_{k,t} = \frac{n_k^{(t)} + \beta_t}{\sum_{t=1}^{V} n_k^{(t)} + \beta_t} \tag{A.12}$$

$$\theta_{m,k} = \frac{n_m^{(k)} + \alpha_t}{\sum_{k=1}^{K} n_m^{(k)} + \alpha_k} \tag{A.13}$$

We remark that $\boldsymbol{\alpha}$ ($\boldsymbol{\beta}$) control the prior distribution for the word-per-topic (topic-per-article) distributions. Generally, a higher (lower) value of $\boldsymbol{\alpha}$ results in each topic distribution having many (few) large-density words. Likewise, a higher (lower) value of $\boldsymbol{\beta}$ results in each article being a mixture of many (few) topics. We use symmetric Dirichlet, meaning that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ reduce to scalars and we set them to their commonly-employed values $\alpha = 1/K$ and $\beta = 1/10$. Table A.1 provides an overview of the introduced variables in this section.

[Table 9 about here.]

B Textual time series: A Chronology of climate-related releases

In this Appendix, we provide a chronology of climate related news releases reflected by the spikes in each one of our textual factors.

B.1 Natural Disasters

November 2000: Rainfall in Southeast Asia and the time duration of drought across Central Asia, reached record-highs over the previous 100 years. At the same time, large parts of Europe also experienced severe floods, and Britain in particular suffered the worst flood in 50 years.

July 2001: Chinese authorities plan a 300-metre-high Xiaowan dam, to help relieve the heavy annual flooding in the Mekong river delta, which has become more frequent and intense over the years.

January 2006: Extreme cold winter snap that affected all of Europe, from Moscow to Paris and caused hundreds of deaths.

Early 2007: A series of record-breaking weather events, ranging from flooding in Asia to heatwaves in Europe and snowfall in South Africa.

August 2007: Hurricane Dean, a category-5 hurricane with a power comparable to Katrina, battered the Caribbean. At the same time, Sahel Africa and South Asia were devastated by floods, Britain suffered the worst flood in 60 years, and Turkey and Australia a pronounced drought.

August 2008: Eastern India suffered its worst flood in 50 years, destroying 250,000 houses and affecting about two million people. In that same month, the melting of arctic ice due to record-high temperature caused floods also in Canada, whereas Cyprus suffered its worst ever drought.

February 2009: Wildfires in Australia, causing hundreds of deaths, and on the heavy rains and floods that followed one week after the fire was put under control.

December 2009: In parallel with the Copenhagen conference on climate change, news report on the increased incidence of natural disasters around the globe, calling for urgent international cooperation.

January 2011: Floods in Australia extensively covered by media.

February 2012: News mostly reported on cyclone Yasi in Australia, and on a report by the Asia Development Bank, which warned about the risk of mass migration linked to the increased occurrence of natural disasters in the region.

March 2015: Cyclone Pam, the second most intense tropical cyclone of the South Pacific Ocean in terms of sustained wind, inflicted one of the worst natural disasters to the Pacific island of Vanuatu, over its history. At the same time, Chile and Zimbabwe suffered heavy floods. In March 2015, news also report extensively on the third United Nations (U.N.) conference on Disaster Risk Reduction; U.N. member States met to set a common policy framework to deal with the catastrophic consequences of natural disasters.

November 2015: Wildfires raged over Southern Australia, while Beijing and New Delhi were covered by a choking cloud of pollution, forcing inhabitants of the Chinese capital to stay indoors.

November 2018: Wildfires raged in South California, destroying about 2,000 homes and led more than 500,000 civilians to evacuate their homes, while Hurricane Paloma battered the British Caribbean.

B.2 Global Warming

February 2007: Publication of the IPCC report, a U.N. organization that groups 2,500 researchers from more than 130 nations. For the first time, the report attributed climate change to human actions with a probability of 90%. This was a substantial upward revision with respect to previous publications, which also implied potentially catastrophic scenarios for the end of the century.

April 2007: IPCC outlined the likely impacts of warming and noticed that rising temperatures could lead to more hunger, water shortage, more extinctions of animals and plants, crop yields could drop by 50% by 2020 in some countries, and projected a steadily shrinking of the arctic sea ice in summers. It also stated that by the 2080s, millions of people will be threatened by floods because of rising sea levels, especially around river deltas in Asia and Africa and on small islands.

November 2007: IPCC agreed to a set of guidelines for policymakers to cope with the rising risks of climate change, urging for prompt actions to reduce drastically greenhouse gas emissions.

December 2009: News reflected the coordinated attempt of the British Meteorological Office and the U.N. Panel on Climate Change to reiterate the validity of scientific evidence on human's actions causing climate change. This followed accusations by climate change sceptics who seized leaked emails from the University of East Anglia and accused climate experts of colluding to manipulate data.

November 2015: A number of articles discussed the World Meteorological Organization announcement that 2015 was the hottest year ever, and that temperatures in 2015 were likely to reach the milestone of 1 degree Celsius above the pre-industrial era.

B.3 International Summits

November 2000: The Hague meeting on climate change. The meeting took place to ratify (i.e. make it legally binding) the Kyoto protocol of 1997, in which countries expressed their joint intention to reduce greenhouse gases by an average of 5% by 2008-2012. In Hague, countries discussed the concept of "emission trading", which would allow companies to buy and sell the right to pollute. The countries failed to ratify the Kyoto agreement, yet they took a first step in that direction.

July 2001: Bonn meeting. This continued the negotiations started in Hague, yet no ratification of the Kyoto protocol was achieved either.

February 2005: Ratification of Kyoto protocol. U.S. did not agree, as President Bush decided to refrain. Even though U.S. did not ratify the Kyoto protocol at the federal level, a number of States on the east and west coasts began to set up regional climate pacts that would require power companies to trade emissions of heat-trapping gases, moving de facto U.S. climate change policy more in line with the aim of the international treaties.

May 2006: First transaction in the Chicago Climate Exchange linking greenhouse gas emission trading systems in Europe and North America.

February 2007: The Global Legislators Organisation held a meeting of the G8+5 (the five leading emerging economies: Brazil, China, India, Mexico, and South Africa) Climate Change Dialogue, where a non-binding agreement was reached to cooperate on tackling global warming. The group accepted that there should be a global rule on emission caps and on trading carbon emissions schemes, applying to both industrialized nations and developing countries. The group hoped this policy to be in place by 2009, to supersede the Kyoto Protocol.

December 2007: Delegates from more than 180 nations met in Bali to start negotiations on a new climate change treaty to succeed the Kyoto Protocol.

December 2008: U.N. Climate change Conference in Poznan, continuing previous negotiations, in preparation for the Copenhagen Summit of December 2009.

June 2009: U.N. Climate change Conference in Bonn, continuing previous negotiations, in preparation for the Copenhagen Summit of December 2009.

December 2009: Copenhagen Summit. The Copenhagen accord declared that climate change is one of the greatest challenges nowadays, and that actions should be taken to ensure that temperature would not increase beyond 2 degrees Celsius. However, the document was not legally binding and did not contain any legally binding commitments for reducing CO_2-emissions, only an intention to reduce carbon emissions further.

November 2012: U.N. climate change conference in Doha.

B.4 U.S. climate policy

November 2006: The Democratic party takes control of the House of Representatives, and puts pressure on capping carbon emissions, despite the opposition of President Bush.

January 2007: Press coverage reflects the climate related content in the Bush's State of the Union Address. Bush called for doubling the capacity of the strategic petroleum reserve and for an increase in transportation fuel standards, but did not advocate limits on the emission of greenhouse gases.

June 2007: An environmental funding bill is passed in the House of Representatives, specifying for the first time a cap on greenhouse emissions.

June 2008: The Lieberman-Warner Climate Security Act reaches the Senate floor, initiating a debate on comprehensive climate change legislation.

January 2009: Obama takes office, setting the stage for reversing the lack of attention to climate change issues that characterized the Bush administration.

September 2009: The House of Representative passes the first comprehensive climate change bill, promoting the use of clean energy sources to suppress the use of fossil fuels.

April 2010: The BP oil spill in the Gulf of Mexico attracted vast media coverage. The political consequence was to upset hopes for winning bipartisan support to U.S. climate legislation, which rested on including measures to encourage more off-shore drilling, that were key to attract support from Republicans.

November 2010: Republicans took back control of the House of Representatives and gained seats in the Senate in the off-year elections. This decreased chances that the U.S. congress would pass a climate bill with substantial reforms, during President Obama's first term.

March 2011: Republicans in the U.S. House of Representatives introduced a bill that would permanently stop the environmental protection agency from regulating emissions blamed for warming the planet.

November 2012: Obama is confirmed president of the U.S. for another term, but Republicans confirm control over the House of Representatives.

February 2013: Obama's State of the Union Speech. He confirms again his commitment to fight climate change.

November 2014: The Democratic party loses control of the Senate in the mid-term elections.

January 2015: Republican Senators introduced a bill to approve the keystone XL pipeline, a major infrastructure for transporting oil from Canada to Texas, despite Obama's opposition.

November 2016: Donald Trump wins the elections, wowing to undo whatever progress Obama was able to make. In the first few months following his election, the news often report his claim that climate change is a hoax.

December 2016: Trump nominates Scott Pruitt to lead the Environment Protection Agency.

June 2017: Trump officially declares that the U.S. would abandon the Paris agreement.

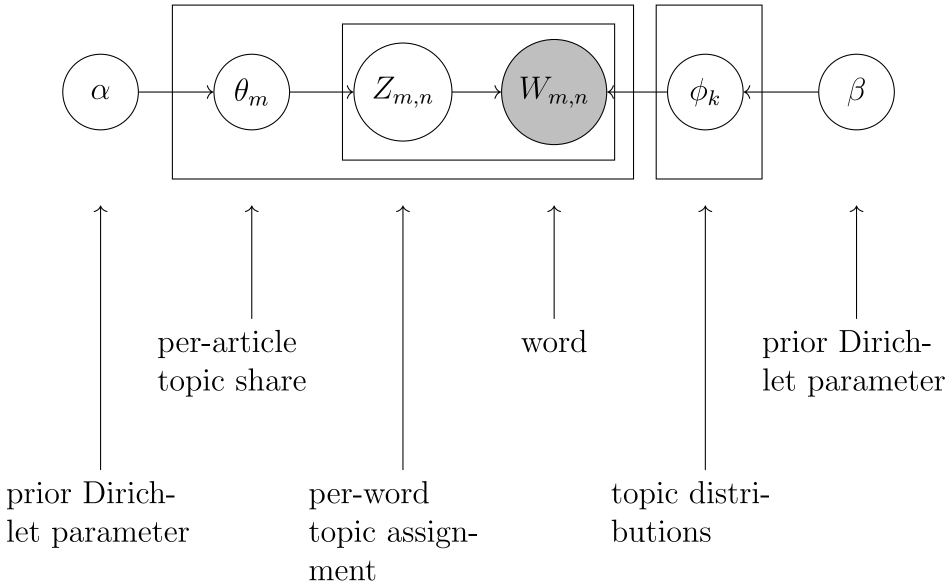


Figure A.1. LDA visualized with plate notation (adapted from Blei et al. (2003)).

Symbol	Description
K	Number of topics
V	Number of words in the vocabulary
M	Number of articles
N_m	Number of words in article m
$\boldsymbol{\alpha}$	Positive V-vector
$\boldsymbol{\beta}$	Positive K-vector
$\boldsymbol{\phi}_k$	Distribution over words for topic k (positive V-vector)
$\boldsymbol{\theta}_m$	Distribution over topics for article m (positive K-vector)
$z_{m,n}$	Topic assigned to word n of article m
$w_{m,n}$	Word n of article m
$n_m^{(k)}$	Number of word tokens in the mth article assigned to the kth topic
$n_k^{(t)}$	Number of times the t term in the vocabulary has been assigned the kth topic

Table A.1. **Notation of LDA**.