Hedging Climate Change News (Replicate)

Reference: Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. The Review of Financial Studies, 33(3), 1184-1216.

Shih-Lun Huang Yu-Chien Li Pi-Wen Lee

I. Summary

The research purpose of this paper is to construct a portfolio that can hedge the climate change risk. To achieve this goal, first, the authors extract innovations of climate news series through textual analysis of newspapers and construct a Climate Change News Index. Second, they use this index to construct a proxy for climate change, called CC_t, and use E-Score to construct a proxy for climate risk exposure. Finally, using all the variables calculated in the previous steps to select stocks and calculate their weights to form the hedge portfolios. Various approaches to construct such hedge portfolios have been proposed in the literature. The two main ones are cross-sectional regressions like Fama-MacBeth and mimicking portfolio approach.

In the result, the research purpose is to check if the return of the hedge portfolios we construct becomes higher as the Climate Change News Index gets higher. If so, investors can use it to hedge the climate change risk.

II. Research Methods and Procedures

1. Construct WSJ Climate News Index

1.1 Climate Change Vocabulary Word Cloud

To start with, we collect 19 climate change white papers and 55 climate change glossaries to construct a corpus of climate change vocabulary. Since 11 of the glossaries are missing, we only have 44 climate change glossaries.

To plot a word cloud that displays the important climate change vocabulary from all the authoritative climate change texts, we aggregate all the terms into a single document and do the following text processing steps.

1. Cut all the words into unigrams and bigrams.
2. Remove stop-words and digits.
3. Do lemmatization.
4. Count the frequency of each term.

After these four procedures, we plot a word cloud in figure(1) with term sizes proportional to their frequency.

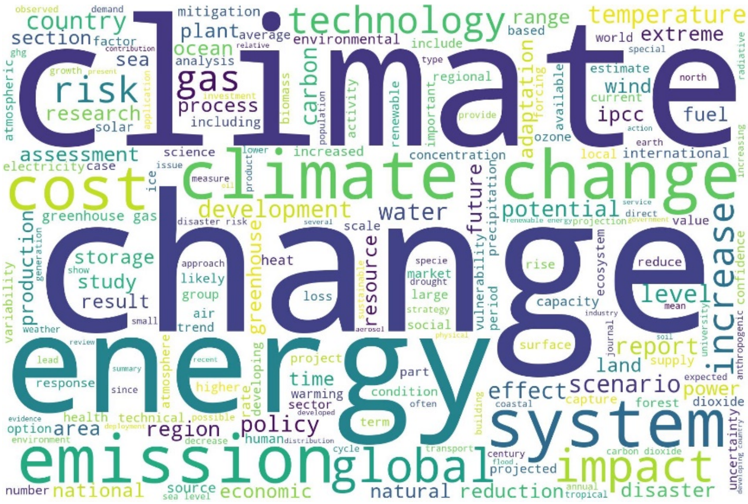

Figure 1. Word cloud result

Comparing with the figure from the paper, we notice that the important words about climate change such as "climate", "change", "climate change", and "emission" are well captured. However, we also notice some differences between the two. For example, "carbon" and "greenhouse" are large in the paper, but ours are much smaller. Moreover, "energy" is large in our result, but it is much smaller in the figure of the paper. The reasonable explanation for the difference is that we lose 11 glossaries, so the frequency of terms might be different.

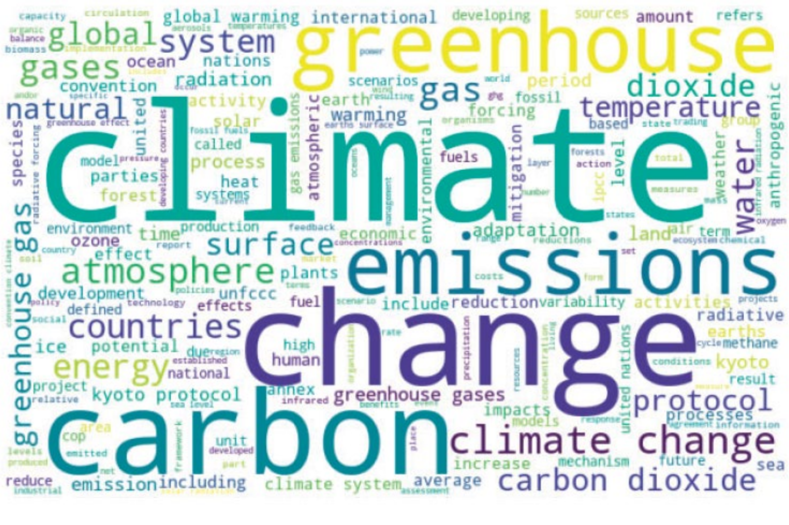

Figure2. Word Cloud from the paper

1.2 Wall Street Journal tf-idf

After collecting the corpus of climate change vocabulary (CCV), we implement web scraping to collect all the Wall Street Journal daily news from the official website and also use ProQuest Dataset to collect some Wall Street Journal articles. Because there are about 200 articles a day, it takes too much time to crawl down all the data from 1984 to 2017 as the paper did. Therefore, we only collect all the daily news from 2013 to June 2, 2016, and arrange all the articles in a time series format as below. Each cell is an article on that day, thus each row represents all the articles on that day.

Date	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2014/03/27	U.S. Defe	The yield	Canada'	PBS' s Y	United Pa	Individual	Permira ai	Johnson &	Washingto	The libera	El Salvado	The Conn	The S&P	WASHIN	Burger ch	PLYMOU	Hollywoo	The highly	The U.S. i	CHICAG
2014/03/28	General M	PHOENIX	Every rece	Dear Dan,	NEW YO	Islanders (	The Long	Investors i	The New	New York	Darth Vad	The Nets'	Dallas Mu	Taliban gu	Willian B	Elaine's at	A plate of	Rookie po	A seagull	OSO, W
2014/03/29	A modera	OSO, Was	KABUL—	RIYADH,	WASHIN	BUENOS	If you hav	LONDON	PERTH, A	A Slovak	LAS VEG	INDIANA	RIYADH,	NEW YO	ANAHEI	The seaso	General M	WASHIN	ISTANBU	BANGK
2014/03/30	Liberals k	ALBANY	After mor	Memories	LOS ANG	The pianis	A British	Since oper	Natural-ga	New York	At the out	Scientists	Early nati	Jeremiah I	TOKYO—	You migh	Greece's F	LONDON	KIEV, Uk	When I
2014/03/31	Penicillin	WASHIN	New York	After year	Hours afte	SVALBA	TOKYO—	Prime Mir	As some c	The U.S. (	The Claim	Sharehold	WASHIN	ALBANY	The harsh	New prob	New York	Comparin	Bicycle ra	The Chi
2014/04/10	RINGGO	Finding ar	Every yea	The resign	Joel and E	Sylvia Ma	Kathleen S	Manhattan	ON A RE	Once used	NEWARK	For some	The cosm	New York	With 44 s	The Justic	RALEIGH	In the wee	Birdland 3	Who co
2014/04/23	RUSTEN	MAYABU	Once an a	A day afte	Greta Ger	Former W	HONG K	A DJ at a	The Metro	Barrick G	Mets man	United Na	DHAKA,	Not every	The forme	In Alain V	HONG K	To unders	The Mets'	The U.S
2014/06/10	BUENOS	Prospect F	BRIDGEF	LONDON	HEIDELB	Shortly af	SEATTLE	MELBOU	American	A Chinese	Donald Tr	Rangers fa	Success A	Somehow,	BRUSSEL	São Paulo	Two south	Jane Menc	The growi	Eight cu
2014/08/15	Mike Wei	In a famou	Guests wh	Republica	The Justic	The Giant	Lousy Inte	The tree-l	For the M	Outfielder	SAY CHE	"Alternati	The use of	Two Braz	Texas Gov	Winged in	Dear Dan,	TEL AVI	New York	A data b
2014/09/04	A New Yo	A federal	When the	New York	Certain pr	BP BP 0.1	TorontoPl	Filmmake	WEST PO	RICHMO	SURREY,	Apple Inc.	WASHIN	His arm a	CARSON	Delaware	The Toron	WASHIN	Last seaso	Despera
2014/11/30	Yes, it' s	New York	After wee	I am told	BEIJING-	TOKYO—	Energy sto	HONG K	Home Equ	What shou	JACKSON	Real-estat	FERGUSO	The premi	BOGOTÁ	Zankel Ha	The Nets	As accepta	Famous w	Two yea
2015/01/01	HONOLU	When the	New York	Economic	I hate mee	Teresa Jor	Mario Cuo	President	New York	When Jets	A former	EAST RU	Three peo	Some of A	Horton Fo	Alliance F	The U.S. (	If "Baby	I' M IN	Broadwa
2015/02/15	MANDAI	SEOUL—	CARTAG	Michele F	A problem	Looking to	STINSON	Highlights	PARIS—	Federal re	BERLIN-	ArcelorMi	BERLIN-	It is a mar	SINGAPO	A Russian	Last week	When lau	South Aus	Islamist
2015/02/22	Last week	While Ne	How do y	These are	Libya' s	David Hu	A blast du	Along the	Chamber	Susan Car	U.S. air-sa	CHICAGO	Cherie Fu	JoAnne K	In 1949, w	For the se	On a snow	Gov. Chri	The stock	About 1
2015/03/01	Aereo Inc.	A homeles	U.S. bond	A conserv	BAGHDA	BARCEL	As Ameri	Zankel Ha	STANSTE	Pharmace	TOKYO—	My marria	It was yea	The Rang	Warren B	For 50 yea	Amgen In	A dream c	Index fund	Federal
2015/03/08	NEW JER	An article	Developm	The New	It was a	The city n	Foxconn,	With the s	BRASÍLI	If you' re	Technolog	General M	Money-ma	RFR Real	Jets runni	As Ameri	The ranks	Blackston	Television	Althoug
2015/03/15	Have you	China' s	Clovis Cla	It' s an i	Age: 76H	CARACA	Some of u	I have rea	ALBANY	$308.3 bil	RETIREM	It isn' t j	Ken Pybu	HERAT,	BOSTON	The learni	At the Bo	Blackston	"You wo	Sweepin
2015/03/22	In Februar	The push	Fee-for-se	The battle	When I m	Some of o	Darden Re	SINGAPO	For years,	For the pa	First, a bit	Starbucks	A major s	Decades b	Military v	Lee Kuan	Prospects	During Ne	'Opus P	A federa
2015/03/29	After year	TIKRIT, I	What do t	A new fas	HOUSTO	Saudi war	TRENTO	Rangers g	SYDNEY	NEW DE	If the Nets	One night	SYDNEY	As negotia	Investigat	ALBANY	GLENDA	DUFRYS	Indiana G	U.S. far
2015/04/05	A suspect	The Apple	The perce	The unrele	SubCultur	The iron-c	The winne	WASHIN	Derek Jete	EGYPTA	Preet Bha	Samsung I	INDIANA	Visits to F	The world	On many	The dollar	This shoul	Storm clo	Greek P
2015/04/12	Generous	The first-r	Entering S	Charities a	Do home	New York	There is o	Wells Far	I' m fairl	Until Sun	When wea	The vendo	Four new	Closed-en	PANAMA	Lots of re	When you	New York	Digital dis	SINGAP

Table 1. Wall Street Journal daily news data

We will merge all the articles in each row into a single document before computing the tf-idf. To be more specific, it doesn't mean that we merge the texts in a real document file. It means that we consider all the articles in a row (each daily data) as a different document while computing the inverse document frequency (idf). And next, we have to do the text processing procedures as mentioned before. But because of the memory limitation of our computer, we only cut the word into unigrams. Finally, we transform the terms of WSJ daily news into the term frequency-inverse document frequency (tf-idf) matrix. The following is the formula for tf-idf.

TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

1.3 Compute Cosine Similarity

For the climate change vocabulary (CCV), we've got the term frequency (tf) before drawing the word cloud. And because CCV is aggregated into a single document, so when computing the inverse document frequency (idf), we have to apply it from the calculation of WSJ corpus and eventually get the tf-idf matrix for CCV.

After transforming WSJ and CCV into the tf-idf matrix, we compute the "cosine similarity" between the tf-idf matrix of CCV and each daily WSJ, which reveals the similarity between the two vectors. Therefore, the day that WSJ uses the same terms in the same proportion as the CCV will earn a cosine similarity value of one, and the day with no words about climate earns a value of zero. Besides, the authors use cosine similarity rather than traditional distance measures because term-frequency vectors are typically very long and sparse (i.e., they have many 0 values). If we use the traditional distance measures for such sparse numeric data, two term-frequency vectors may have many 0 values in common, meaning that the corresponding documents are similar (but not that similar). However, cosine similarity measures the similarity between two vectors of an inner product space (i.e., measured by the cosine of the angle between two vectors). Thus, it is a better way to measure document similarity in a sparse matrix.

In addition, in the paper, the authors scale this similarity value by a factor of 10,000 and finally, we will get the WSJ Climate Change News Index, which describes the fraction of the WSJ dedicated to the topic of climate change each day.

After acquiring the daily WSJ Climate News Index, we translate the monthly value by computing the mean of all daily values in the same month. We show the result of the monthly WSJ Climate News Index in figure(3). When compared with the index figure from the paper, we noticed great differences between the two. To begin with, the index value we obtain mostly lies between 1500-1900, whereas the paper is nowhere above 200. Secondly, the line we plot is smoother overall, while the line from the paper is jagged, and can see obvious peaks. What caught our attention the most is the drastic drop around June of 2015. To make sense of the cause behind this issue, we plot the daily WSJ Climate News Index in figure(5) and the daily amount of WSJ news in figure(6). The number of news during June of 2015 appears to be less than others, which got us wondering whether there exists a relationship between the Index and the amount of news collected. However, after computing the correlation of the two, we got a result of -0.08, which does not seem to stand by our assumption.

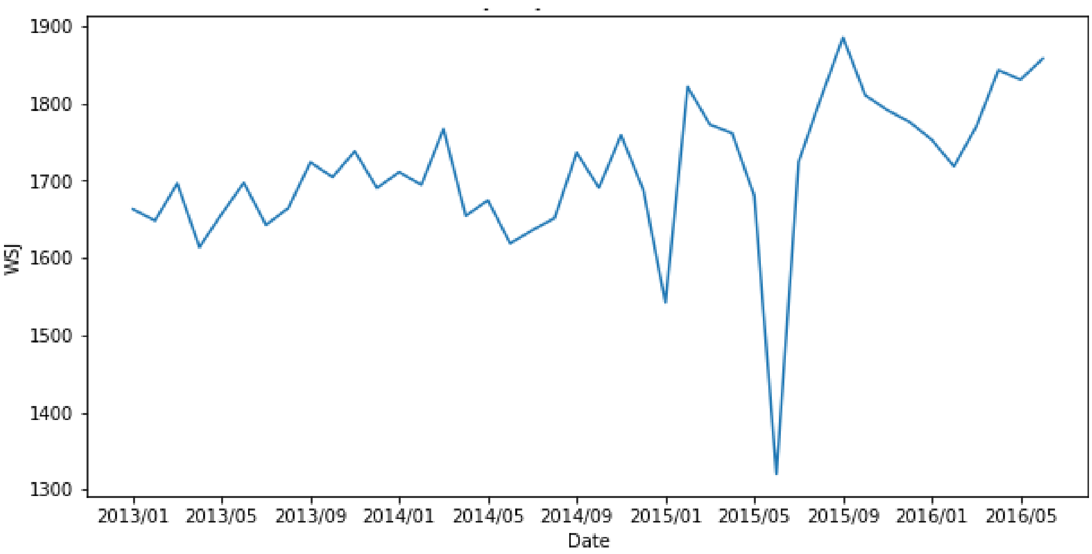

Figure 3. Monthly WSJ Climate Change Index

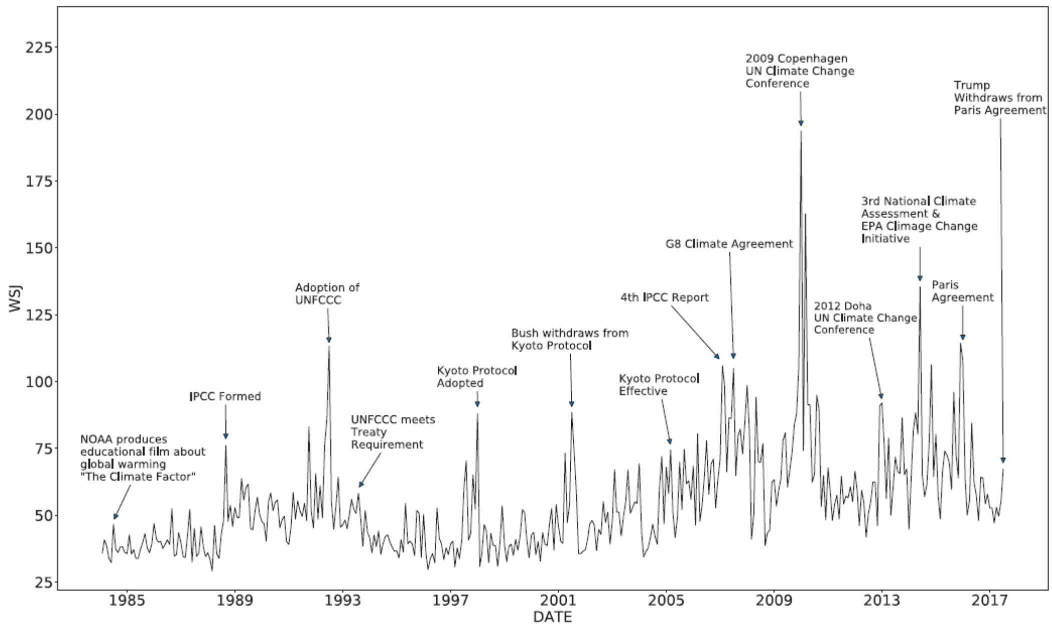

Figure 4. Monthly WSJ Climate Change Index from the paper

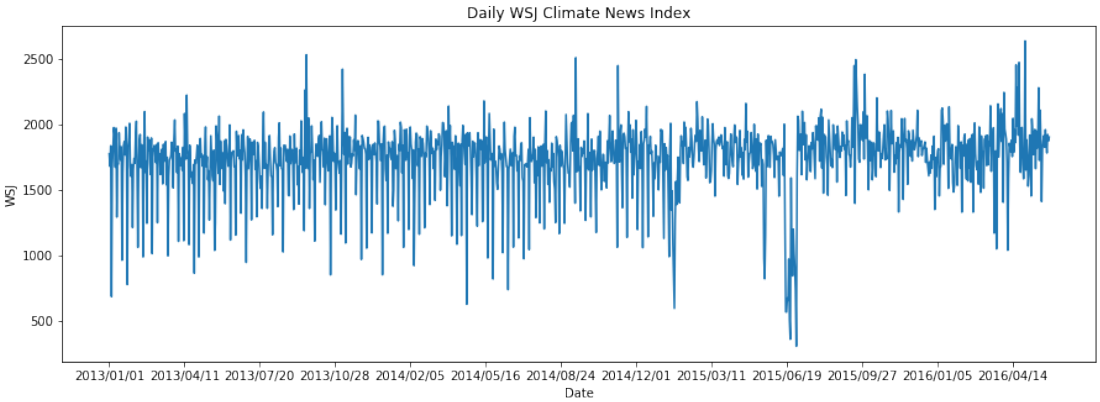

Figure 5. Daily WSJ Climate Change Index

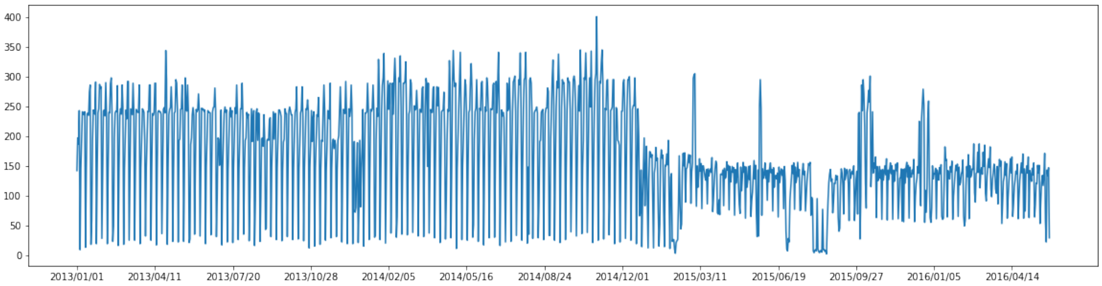

Figure 6. Daily WSJ news count

We also suspect that the great difference between the trend of our result and the paper's may be ascribed to the scale of the index. By standardizing our index and the paper's index, obtained from one of the author's website, the result is plotted in figure(7). An interesting finding is that despite still looking unalike overall, the two indexes seem to capture the same trend from 2013 to 2014.

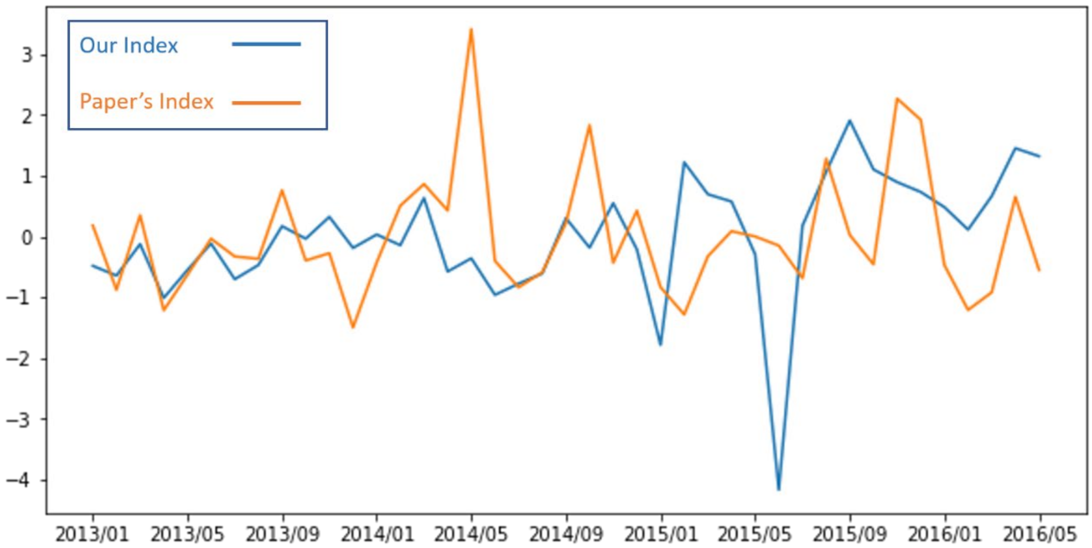

Figure 7. Monthly WSJ Climate Change Index

So, what causes the difference? A possible explanation is that the terms we use to calculate the *tf-idf* only contain unigram. Due to the difficulty of memory limitation, we weren't able to run both unigram and bigram like the research did, which we believe may have resulted in a greater value of our index. Our reason is that by omitting bigram terms we not only excluded a portion of climate change-related words but also simultaneously removed a greater amount of non-climate related terms, thus, causing the value of cosine-similarity to increase. Another factor may be that the measure used to translate the daily WSJ climate change index to the monthly climate change index is different. We obtained our monthly value by calculating the mean value of the daily index in the same month; however, the authors did not mention how they did it.

2. Construct proxy for innovations in climate related news, CC_t

Due to the lack of data for the monthly WSJ climate change index, with only 42 records, for the remainder of our project, we use the daily WSJ Climate Change Index.

The value of CC_t is the residual acquired from performing an autoregressive proc,ess with WSJ climate change news index. First we run the augmented Dickey-Fuller (ADF) test to ensure the stationarity of our data, then we capture the order of AR model needed for our data, and finally run the model to acquire daily residual. The result shows that our data is indeed stationary, with a -value of $3.1842 * 10^{-7}$. We then plot the graph of partial autocorrelation in figure(8) to capture the necessary lag and noticed that it requires a model with a lag of three.

Thus, an AR(3) was conducted, and with the residual of the model, we construct the daily CC_t as a proxy for innovations in climate related news.

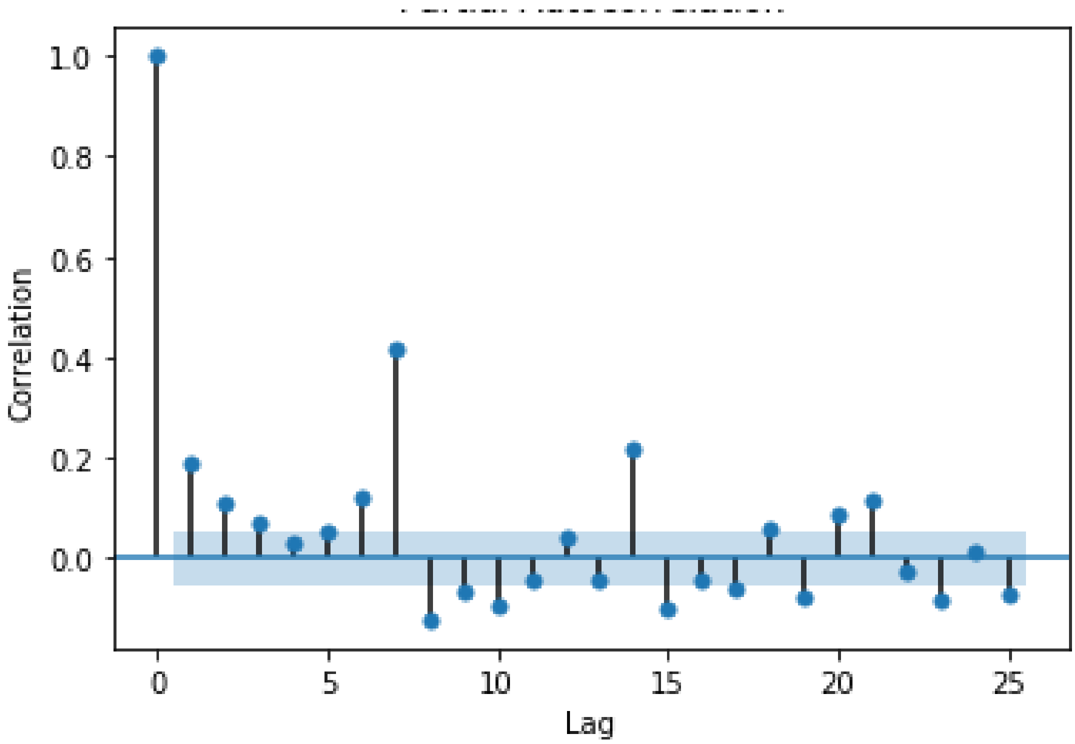

Figure 8. Partial Autocorrelation

AutoReg Model Results

Dep. Variable:	y	No. Observations:	1250
Model:	AutoReg(3)	Log Likelihood	-8709.927
Method:	Conditional MLE	S.D. of innovations	261.325
Date:	Mon, 14 Jun 2021	AIC	11.140
Time:	15:01:28	BIC	11.160
Sample:	3	HQIC	11.147
	1250		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	1164.9478	71.926	16.196	0.000	1023.975	1305.920
y.L1	0.1602	0.028	5.670	0.000	0.105	0.216
y.L2	0.0968	0.028	3.398	0.001	0.041	0.153
y.L3	0.0721	0.028	2.553	0.011	0.017	0.127

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.7813	-0.0000j	1.7813	-0.0000
AR.2	-1.5616	-2.3121j	2.7900	-0.3445
AR.3	-1.5616	+2.3121j	2.7900	0.3445

Table 2. AR(3) result

3. Estimate climate risk exposure for individual stocks

3.1 Theoretical Measure

For the final part of our project, we use the daily CC_t to distinguish the stocks for climate risk hedging portfolio. Equation(1) is the model used to obtain climate risk exposure β_{cc}, which then is used to select the stocks for our the portfolio. As mentioned in paper, the portfolio is constructed with a mimic approach, where the authors select stocks that have the greatest climate risk exposure to capture climate change innovations. Thus, our goal will be to pick out the stocks with relatively large absolute value of β_{cc}. To run this regression model, we will need to first obtain v_t ,which is other risk factors, γ_{cc} and γ which represent risk premium for climate news factors and other risk factors correspondingly.

$$\underbrace{r_t}_{n\times 1}=(\underbrace{\beta_{CC}}_{n\times 1}\underbrace{\gamma_{CC}}_{1\times 1}+\underbrace{\beta_{CC}}_{n\times 1}\underbrace{(CC_t-E[CC_t])}_{1\times 1})+(\underbrace{\beta}_{n\times p}\underbrace{\gamma}_{p\times 1}+\underbrace{\beta}_{n\times p}\underbrace{v_t}_{p\times 1})+\underbrace{u_t}_{n\times 1}.$$

Equation 1

We denote by r_t an $n\times 1$ vector of excess returns over the risk-free rate of n assets at time t . We assume that these returns follow a linear factor model, in which asset returns are driven by innovations in climate news.

To handle these variables, we follow Fama-MacBeth's cross-sectional regression approach[1]. First, we apply equation(2) the Fama-French three-factors model with an additional factor of CC to estimate β_{cc} and β from the time-series regression, equation(3). Then we input these betas into the second cross-sectional regression of return r_t to obtain h_t^{CC} and h_t, the returns of the hedge portfolios. Finally, by computing the time-series means of h_t^{CC} and h_t we can recover the risk premiums, γ_{cc} and γ.

3.2 Empirical Results

Apply the Fama-MacBeth procedure as the former passage mentioned, we need to take a stand on all the factors in the model: CC_t and v_t. Since CC_t has been constructed ahead. Next, we need to obtain v_t ,which is other risk factors as papers said. There is no detail information what v_t $(p \times 1)$ factors are, therefore we view v_t as Fama-French three-factors, which makes sense at all. We first obtained the Fama-French three-factors data daily from *Kenneth R. French's website* [2]. We crawl all the files with Python rather than download the csv file, which provides immediate update if we want

[1] https://en.wikipedia.org/wiki/Fama%E2%80%93MacBeth_regression
[2] Kenneth R. French - https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

to run the regression to today.[3] The 3 factors data show in table(3). We take only NASDAQ all traded securities as our regression data, although the paper includes only common equity securities for firms traded on the NYSE, AMEX and NASDAQ. Following Amihud (2002) and many others, and then excludes penny stocks. We should have done better here. We first download all time Nasdaq traded symbol, and then crawl all the return during Jan.2014 to June.2016, and then delete all the return[4] contains NaN, which represents that the symbol didn't exist during that time. We have *4023 securities* during the time. And we merge the CC which is the AR(3) residual and υ_t to this format.

	Date	CC	RMRF	SMB	HML	RF	A-Rf	ACGL-Rf	ADRE-Rf	AEFC-Rf	...	BXP-Rf	BXS-Rf	BYD-Rf	BYFC-Rf	BYM-Rf
0	2014-01-06	-550.530446	-0.0034	-0.0058	0.0028	0.00000	-0.004919	-0.020516	-0.026689	-0.014121	...	0.011123	-0.010454	-0.008734	0.060606	0.005439
1	2014-01-07	230.351345	0.0068	0.0039	-0.0039	0.00000	0.014301	-0.012002	-0.000851	0.025489	...	0.006425	0.006095	0.039648	0.019048	0.006182
2	2014-01-08	95.659715	0.0004	0.0001	-0.0011	0.00000	0.016362	0.010972	-0.009463	0.006233	...	-0.006674	0.000404	0.010170	0.009346	0.003073
3	2014-01-09	100.407901	0.0002	0.0020	-0.0042	0.00000	0.000342	0.003799	-0.028184	0.000522	...	0.000487	0.000404	-0.005872	0.009259	-0.003063
4	2014-01-10	-12.440586	0.0027	0.0055	-0.0082	0.00000	0.008902	-0.002741	0.019465	0.008503	...	0.016547	0.006457	0.037131	0.100917	0.007680
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
602	2016-05-26	173.745975	-0.0004	-0.0009	-0.0053	0.00001	0.001971	-0.003722	-0.018915	-0.009635	...	-0.003682	-0.015830	-0.005404	-0.000010	0.008476
603	2016-05-27	44.310655	0.0049	0.0033	-0.0017	0.00001	0.006802	0.014185	0.023288	0.015175	...	0.004878	0.013103	0.020597	-0.000010	0.001284
604	2016-05-31	152.687324	-0.0001	0.0050	-0.0034	0.00001	0.001518	0.033144	-0.004664	0.003181	...	0.001744	-0.002515	0.004772	-0.026188	-0.005181
605	2016-06-01	95.063068	0.0020	0.0065	-0.0021	0.00001	0.001298	0.001925	0.026245	0.011918	...	0.011771	0.002083	0.026960	-0.032268	0.009737
606	2016-06-02	123.711661	0.0035	0.0031	-0.0055	0.00001	-0.001098	0.008390	-0.015991	0.026644	...	0.001170	-0.000428	0.014923	0.027768	-0.005158

607 rows × 4029 columns

Table 3. Nasdaq stock returns and CC (partial)

As the Appendix A.1 Review of the Fama-MacBeth Approach shows, we can run regression with the factors: CC_t and υ_t. First, we apply the Fama-French three-factors model with an additional factor of CCt to estimate βcc and β from the time-series regression, equation(3).

$$\upsilon_t: \qquad r^i_{t,v} = \beta^i_1(R_m - R_f) + \beta^i_2 \cdot SMB + \beta^i_3 \cdot HML + \alpha \tag{2}$$

$$r^i_t = \alpha^i + \beta^i_{CC} CC_t + \beta^i v_t + \mu_t \tag{3}$$

[3] Fama-French three factors crawl code provides in Github.

[4] Return is crawled from yahoo finance. https://finance.yahoo.com

	A-Rf	ACGL-Rf	ADRE-Rf	AEFC-Rf	AFIN-Rf	ALRS-Rf	ALUS=-Rf	AMAOW-Rf	AMWD-Rf	ANAT-Rf	...	BXP-Rf	BXS-Rf	BYD-Rf	BYI
beta1	1.19958	1.16838	0.836717	1.53975	1.19144	0.750919	0.990509	1.26833	0.655619	1.29407	...	0.779706	1.22571	1.41379	-0.041
beta2	0.0857505	0.110748	0.233311	0.133527	-0.1899	-0.149593	0.0692709	0.236185	0.832465	0.295015	...	-0.298311	0.976696	0.600208	-0.16
beta3	-0.191504	-0.380735	-0.928985	-1.57167	-0.85745	-0.345385	0.672102	-1.44051	-0.117897	-0.321949	...	-0.193853	1.15056	-0.351245	-1.5
betacc	-4.37257e-07	-5.8984e-07	3.81066e-07	-5.18117e-06	-2.54271e-06	5.35538e-07	-1.7329e-06	-2.80409e-06	-5.91782e-07	5.86046e-07	...	9.66052e-07	2.19534e-06	7.0811e-07	-8.98

4 rows × 4023 columns

Table 4. Beta values of each stock

After the regression by Fama-MacBeth Approach, we obtain a *4x4023* frame which contains 4023 securities' β. In the second step, in each period t , hedge portfolios for all factors are obtained via cross-sectional regressions of returns r_t onto the estimated betas ($\hat{\beta}_{CC}, \hat{\beta}$):

$$r_t = h_t^{cc}\hat{\beta}_{CC} + h_t\hat{\beta} + e_t \qquad (4)$$

With all the variables for our main regression model prepared, we can easily obtain (h_t^{cc}, h_t) by running the regression in equation. Where betas ($\hat{\beta}_{CC}, \hat{\beta}$) are estimated in the former step. The slopes of this regression in each period *t* are precisely the returns of the hedge portfolio in period *t* : h_t^{cc}(that hedges CC_t) and h_t(that hedges the remaining factors υ_t). Their time-series means (the expected excess returns of the hedge portfolios) recover the risk premiums of the factors:

$$E[\, h_t^{cc}] = \gamma_{CC} \quad and \quad E[h_t] = \gamma$$

	h_t^cc	Gamma	A-Rf	ACGL-Rf	ADRE-Rf	AEFC-Rf	AFIN-Rf	ALRS-Rf	ALUS=-Rf	AMAOW-Rf	...	BXP-Rf	BXS-Rf	BYD-Rf
0	B_RMRF	5.175659	1.199578e+00	1.168377e+00	8.367174e-01	1.539745	1.191437	7.509195e-01	0.990509	1.268332	...	7.797056e-01	1.225706	1.413789e+00
1	B_SMB	-5.778870	8.575053e-02	1.107475e-01	2.333112e-01	0.133527	-0.189900	-1.495934e-01	0.069271	0.236185	...	-2.983108e-01	0.976696	6.002083e-01
2	B_HML	-4.101309	-1.915038e-01	-3.807353e-01	-9.289846e-01	-1.571674	-0.857450	-3.453851e-01	0.672102	-1.440509	...	-1.938533e-01	1.150557	-3.512447e-01
3	B_CC	0.000024	-4.372572e-07	-5.898399e-07	3.810656e-07	-0.000005	-0.000003	5.355376e-07	-0.000002	-0.000003	...	9.660518e-07	0.000002	7.081098e-07

4 rows × 4025 columns

Table 5. Beta values and γ of each stock

We can finally take γ_{CC} and γ back into the equation(1). To make it easy calculating, we combine the factor γ_{CC} and *Cfactor*(CC_t - $E[CC_t]$) as one variable, and HMLG($HML + \gamma$), SMBG ($SMB + \gamma$), RMRFG ($R_m - R_f + \gamma$). And run the final regression to get the 4 Beta with 4023 securities.

	Date	HMLG	SMBG	RMRFG	Cfactor	A-Rf	ACGL-Rf	ADRE-Rf	AEFC-Rf	AFIN-Rf	...	BXP-Rf	BXS-Rf	BYD-Rf	BYFC-Rf	BY
0	2014-01-06	-4.098509	-5.78467	5.172259	-553.279811	-0.004919	-0.020516	-0.026689	-0.014121	-0.007088	...	0.011123	-0.010454	-0.008734	0.060606	0.00
1	2014-01-07	-4.105209	-5.77497	5.182459	227.601979	0.014301	-0.012002	-0.000851	0.025489	0.011178	...	0.006425	0.006095	0.039648	0.019048	0.00
2	2014-01-08	-4.102409	-5.77877	5.176059	92.910350	0.016362	0.010972	-0.009463	0.006233	0.009773	...	-0.006674	0.000404	0.010170	0.009346	0.00
3	2014-01-09	-4.105509	-5.77687	5.175859	97.658536	0.000342	0.003799	-0.028184	0.000522	-0.002264	...	0.000487	0.000404	-0.005872	0.009259	-0.00
4	2014-01-10	-4.109509	-5.77337	5.178359	-15.189951	0.008902	-0.002741	0.019465	0.008503	-0.008354	...	0.016547	0.006457	0.037131	0.100917	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
602	2016-05-26	-4.106609	-5.77977	5.175259	170.996610	0.001971	-0.003722	-0.018915	-0.009635	0.009251	...	-0.003682	-0.015830	-0.005404	-0.000010	0.00
603	2016-05-27	-4.103009	-5.77557	5.180559	41.561290	0.006802	0.014185	0.023288	0.015175	-0.003745	...	0.004878	0.013103	0.020597	-0.000010	0.00
604	2016-05-31	-4.104709	-5.77387	5.175559	149.937959	0.001518	0.033144	-0.004664	0.003181	0.014802	...	0.001744	-0.002515	0.004772	-0.026188	-0.00
605	2016-06-01	-4.103409	-5.77237	5.177659	92.313703	0.001298	0.001925	0.026245	0.011918	-0.004645	...	0.011771	0.002083	0.026960	-0.032268	0.00
606	2016-06-02	-4.106809	-5.77577	5.179159	120.962295	-0.001098	0.008390	-0.015991	0.026644	0.012222	...	0.001170	-0.000428	0.014923	0.027768	-0.00

607 rows × 4028 columns

Table 6. Variables for equation (1)

	Beta	A-Rf	ACGL-Rf	ADRE-Rf	AEFC-Rf	AFIN-Rf	ALRS-Rf	ALUS=-Rf	AMAOW-Rf	AMWD-Rf	...	BXP-Rf	BXS-Rf	BYD-Rf
0	BetaRMRFG	1.039281e+00	9.964722e-01	6.691819e-01	1.203204	0.925535	5.988010e-01	0.941929	0.994346	6.786531e-01	...	0.618036	1.324847	1.283339e+00
1	BetaSMBG	6.913787e-01	7.602329e-01	8.662883e-01	1.405035	0.814724	4.251353e-01	0.252814	1.271349	7.454392e-01	...	0.312504	0.602124	1.093069e+00
2	BetaHMLG	3.373022e-01	1.863648e-01	-3.762988e-01	-0.461453	0.019741	1.564410e-01	0.832364	-0.536653	-1.938842e-01	...	0.339481	0.823498	7.909778e-02
3	Betacc	-3.942970e-07	-5.437686e-07	4.259658e-07	-0.000005	-0.000002	5.763060e-07	-0.000002	-0.000003	-5.979549e-07	...	0.000001	0.000002	7.430708e-07

4 rows × 4024 columns

Table 7. Result of equation (1)

BetaSMBG		BetaRMRFG		BetaHLMG		Betacc	
Mean	1.390554	Mean	3.278065	Mean	2.158695	Mean	2.46E-05
Standard Error	0.860359	Standard Error	2.487747	Standard Error	2.045308	Standard Error	3.08E-05
Median	0.433288	Median	0.805168	Median	0.232159	Median	-5.8E-08
Mode	0.760233	Mode	0.996472	Mode	0.186365	Mode	-5.4E-07
Standard Deviation	54.57008	Standard Deviation	157.7906	Standard Deviation	129.728	Standard Deviation	0.001952
Sample Variance	2977.894	Sample Variance	24897.88	Sample Variance	16829.35	Sample Variance	3.81E-06
Kurtosis	2686.938	Kurtosis	3966.162	Kurtosis	3878.383	Kurtosis	3719.935
Skewness	48.13478	Skewness	62.76534	Skewness	61.61314	Skewness	59.2656
Range	3921.976	Range	10464.49	Range	9121.2	Range	0.145585
Minimum	-818.8328	Minimum	-491.072	Minimum	-968.149	Minimum	-0.02421
Maximum	3103.143	Maximum	9973.422	Maximum	8153.051	Maximum	0.121377
Sum	5594.2	Sum	13187.66	Sum	8684.429	Sum	0.098927
Count	4023	Count	4023	Count	4023	Count	4023
Largest(1)	3103.143	Largest(1)	9973.422	Largest(1)	8153.051	Largest(1)	0.121377
Smallest(1)	-818.8328	Smallest(1)	-491.072	Smallest(1)	-968.149	Smallest(1)	-0.02421
Confidence Level(95.0%)	1.68678	Confidence Level(95.0%)	4.877361	Confidence Level(95.0%)	4.009936	Confidence Level(95.0%)	6.03E-05

Table 8. Summary statistics of betas

III. Discussion

We rank the β_{CC}, and select the top three of the biggest and smallest. The biggest one contains chemical company, special purpose acquisition company, gas and oil company. The smallest one contains equity mutual fund, gas and oil company, REITs. Both includes a gas and oil company, and the smallest one contains mutual fund and there are actually some ETF in the place 4th and 5th. The result tells us there are no specific kind of industry has better ability to diversify the risk exposure to climate change. It could probably varies in same industry.

Biggest	1	2	3
	0.00145974	1.285E-05	8.116E-06
	EMN-Rf	EMPW-Rf	ENLC-Rf
Smallest	1	2	3
	-1.101E-05	-1.057E-05	-1.002E-05
	EOS-Rf	EPSN-Rf	ESBA-Rf

Table 9. Top three firms ranked by β_{CC}

Eastman Chemical Company is an American company primarily involved in the chemical industry. Once a subsidiary of Kodak, today it is an independent global specialty materials company that produces a broad range of advanced materials, chemicals and fibers for everyday purposes; Empower Ltd. operates as a blank check company. The Company aims to acquire one and more businesses and assets, via a merger, capital stock exchange, asset acquisition, stock purchase, and reorganization; EnLink Midstream, LLC provides midstream energy services in the United State. The company is involved in gathering, compressing, treating, processing, transporting, storing, and selling natural gas, gas liquids, crude oil and condensate. Its midstream energy asset network includes approximately 11,900 miles of pipelines; 22 natural gas processing plants; 7 fractionators; barge and rail terminals; product storage facilities; brine disposal wells; and a crude oil trucking fleet. The company was incorporated in 2013 and is based in Dallas, Texas.

Figure 9. Top three companies with the greatest β_{CC}

Eaton Vance Enhanced Equity Income Fund II is a closed-ended equity mutual fund launched and managed by Eaton Vance Management. The fund invests in public equity markets of the United States. It seeks to invest in the stocks of companies operating across diversified sectors. The fund primarily invests in growth stocks of mid-cap and large-cap companies; Epsilon Energy Ltd., a natural gas and oil company, engages in the acquisition, development, gathering, and production of oil and gas reserves in the United States. It operates through Upstream and Gathering System segments. As of December 31, 2020, it had total estimated net proved reserves of 88,658 million cubic feet of natural gas reserves and 371,343 barrels of oil and other liquids. Epsilon Energy Ltd. was incorporated in 2005 and is based in Houston, Texas; Empire State Realty OP, L.P. operates as a subsidiary of Empire State Realty Trust, Inc.

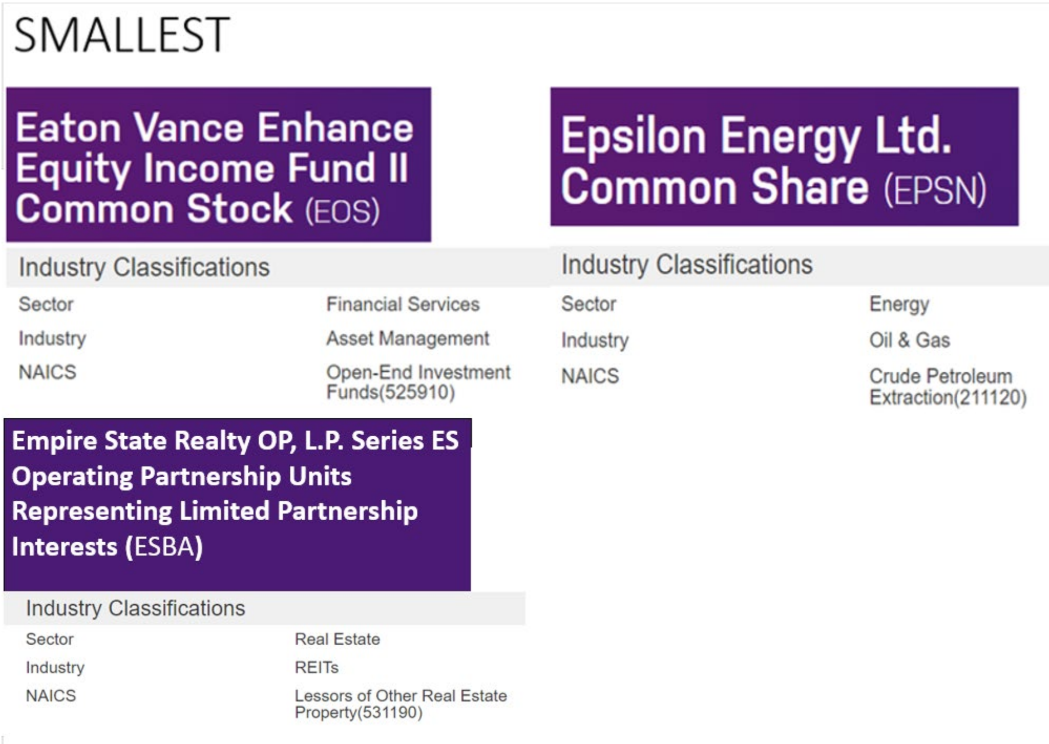

Figure 9. Top three companies with the smallest β_{CC}

To conclude, from the result value of the β_{CC} we obtained, we find it hard to apply it on distinguishing stocks that have great risk exposure to climate innovations. We believe our overall replicate methods and procedure is accurate, however, the main issue that caused our result to deviate from the paper's is data collecting.

To begin with, we missed 11 climate change glossaries when constructing Climate Change Vocabulary. Secondly, not all Wall Street Journal news are acquired through Wall Street Journal archive, news after 2014 were collected from ProQuest. Although we have no understanding of how ProQuest process and select news from Wall Street Journal, from the daily WSJ news count in figure6, we can easily notice a drop in the amount of news starting from 2015 which reveal that ProQuest did screened out some categories of the news from Wall Street Journal archive. In addition, the timeline of our data is not long enough, we only collect 3.5 years out of 34 years of Wall Street Journal news which may make a huge difference in the result of our WSJ Climate Change News Index.

For future research, we should first conduct a more complete data collection. If the result of β_{CC} can significantly convey risk exposure to climate change innovations, we can then carry out the rest of the paper. To continue the research, we will need to collect MSCI KLD scores and Sustainalytics scores to compute the weight of each selected stock in our climate change news portfolio. The result from the paper displays that this portfolio shows prominent effect on tracking climate change innovations when compared with extant ETFs, such as XLE and PBD. Thus, if our portfolio also outperforms these ETFs, then we can confirm its ability to hedge climate change innovations.

IV. Data and Python Code link

GitHub online link:
https://github.com/liyuchien/FRM2021-Group5