

Fractional Integration and Cointegration in Textual Climate Risk Factors

Marina Balboa^a, Ángel León^b, and Antonio Rubia^{*a}

^aDepartment of Financial Economics, University of Alicante

^bDepartment of Economics, University of Alicante

Work in progress. This version: January 2026

Abstract

This paper examines the presence of long memory and fractional cointegration among four textual climate risk factors derived from news-based indices following [Faccini et al. \(2023\)](#): *U.S. climate policy*, *International summits*, *Global warming*, and *Natural disasters*. The results indicate that climate risk factors exhibit non-stationary fractional integration. The estimated degree of persistence is comparable to that observed in realized and implied volatility measures in stock markets. Moreover, we find evidence of fractional cointegration in a system comprising the climate risk factors and the VIX index. This evidence is consistent with the interpretation that textual climate risk factors are partially driven by persistent market-wide sentiment.

Keywords: Fractional cointegration, Long memory, Climate risk, Exact local Whittle.

JEL codes: C32, Q54, G12.

*Corresponding author. We thank Renato Faccini and George Skiadopoulos for generously providing the data analyzed in this paper. Financial support from the Ministerio de Ciencia, Innovación y Universidades through project PID2021-124860NB-I00 is gratefully acknowledged. Any errors are solely the responsibility of the authors.

1 Introduction

Global warming has become a central topic in financial and macroeconomic research; see [Krueger et al. \(2020\)](#) and [Giglio et al. \(2021\)](#) for a review. A major challenge in the analysis of climate risk is the lack of specific data. This problem has motivated the use of textual methods which, on the basis of large corpora of newspaper articles and other news sources, measure the intensity of climate-related topics in the media. This approach generates large samples of regularly-spaced, time-varying textual climate risk factors (TCRF henceforth) which are believed to capture the evolution of investors' sentiment regarding long-run climate hazards. In this spirit, [Engle et al. \(2020\)](#) process articles published in *The Wall Street Journal* to construct a climate-change news-coverage index and discuss how shocks to this index can be dynamically hedged in a portfolio context. The index is computed as the proportion of the texts devoted to issues on climate change each day given keywords in line with the Climate Change Vocabulary. [Faccini et al. \(2023\)](#) apply Latent Dirichlet Allocation (LDA) machine learning to Reuters news to construct climate-risk factors, showing that transition risk is priced in the U.S. stock market; see also, among others, [Wang et al. \(2022\)](#), [Ardia et al. \(2023\)](#), [Bua et al. \(2024\)](#), and [Allahdadi et al. \(2024\)](#).

Given the novelty of this literature, little is formally known about the stochastic nature of TCRF even though this aspect plays a major role in the econometric analysis based on these series. In this paper, we argue that the pattern of serial dependence exhibited by climate risk factors can be accommodated by fractionally integrated (long-memory) models; see [Baillie \(1996\)](#) and [Hualde and Nielsen \(2023\)](#) for a review. Two complementary arguments support this premise. First, textual indices reflect public attention and collective human focus. Within a given topic, major events generate heightened media coverage that decays gradually over time, causing autocorrelation in the resultant measures. Consequently, the long-run of TCRF is plausibly driven by persistent trends of media attention dominated by low-frequency shocks that fractional integration models are well suited to capture. In addition, textual processing aggregates data cross-sectionally. For instance, LDA and other machine learning algorithms construct news-based indices by summing topic probabilities or word counts across thousands of heterogeneous micro-units (news articles). As shown by [Granger \(1980\)](#), cross-sectional aggregation of short-memory processes with heterogeneous characteristics causes hyperbolically decaying autocorrelations, the hallmark of fractional integration. Consequently, long memory in climate textual factors may alternatively arise from cross-sectional aggregation inherent in the construction of these time series.

We study the stochastic nature of TCRF using both univariate fractional integration and fractional cointegration analyses. We focus on the set of textual indices elaborated by [Faccini et al. \(2023\)](#), obtained from textual and narrative analysis of Reuters climate-

change news over January 1st 2000 to January 31st 2025. The core variables in this climate-risk system (jointly referred to as $\mathcal{CR}$ system henceforth) are physical and transition risk measures associated with four topics: (i) U.S. climate policy, (ii) international summits, (iii) global warming, and (iv) natural disasters. To shed light on whether long-range correlation in these series is driven by persistence in media coverage and public sentiment or is caused by LDA cross-aggregation, we examine the long-run dynamics of an augmented system that incorporates VIX, the stock market option-based implied volatility. Realized and option-implied volatility measures are well-known cases of fractionally integrated models. Fractional cointegration in the augmented system (referred to as the $\mathcal{VCR}$ system henceforth) is mainly consistent with the hypothesis of persistence in investors' sentiment because VIX is driven by market-wide uncertainty and investors' risk aversion (Bekaert et al., 2013). In this setting, fractional cointegration is the consequence of common stochastic trends governing the long-term of different proxies for investor sentiment.¹ This analysis provides new insight into the persistence, co-movements, and sentiment-driven dynamics underlying textual climate risk factors.

Our analysis follows in two main stages. First, we discuss the long-memory property of the four climate risk factors and the volatility index from an univariate perspective. We implement semiparametric estimation in the frequency domain using two-staged feasible Exact Local Whittle estimation as advocated by Shimotsu and Phillips (2005) and Shimotsu (2010). This estimation strategy produces consistent and asymptotically normal estimates under general assumptions that remain agnostic about the short-run component of the series. Furthermore, these properties hold regardless of whether the variables under analysis are cointegrated, and regardless of whether fractional integration lies in the stationary or non-stationary region over a wide admissible theoretical range. Owing to these properties, this estimator has been implemented in different papers dealing with long-memory in financial series; see, among others, Berger et al. (2009), Caporin et al. (2013) and Afzal and Sibbertsen (2021).

The evidence from this analysis strongly indicates that both the climate risk factors and the VIX series are governed by fractionally integrated dynamics with a similar order. Given these estimates, we assess the necessary condition for fractional cointegration that the order of fractional integration is common across all variables. To this end, we employ the Wald-type test proposed in Robinson and Yajima (2002) as extended by Nielsen and Shimotsu (2007). The results provide support for the existence of a homogeneous fractional integration coefficient across the series.

¹Cointegration is often interpreted as the existence of long-run equilibria driven by market forces or economic agents. As emphasized by Granger (1986), however, cointegration does not necessarily require economic decision-making; it more generally describes the tendency of a system of variables to evolve toward a restricted region of the outcome space.

This result motivates the interest in exploring fractional cointegration relationships. Consequently, in a subsequent analysis we address the existence of common underlying trends and long-run equilibria in the $\mathcal{VCR}$ system by estimating a Fractional Cointegrated Vector Autoregression (FCVAR) model of Johansen and Nielsen (2010). This analysis not only allows to formally test the existence of common trends through the estimation of the cointegrating rank, but also characterize the cointegrating vectors and the long-term loadings that feature the error correcting mechanism in the system.

The FCVAR results show that climate risk and VIX measures exhibit non-stationary fractional cointegration over the analyzed period with two persistent common stochastic trends. Moreover, the estimated loading coefficients reveal heterogeneous adjustment patterns, shedding light on how each climate risk dimension contributes to restoring long-run equilibrium. This richer characterization of the long-run structure provides new insight into the persistence, co-movements, and sentiment-driven dynamics underlying textual climate risk indicators.

The remaining of this paper is organized as follows. Section 2 presents the data analyzed in this paper. Section 3 discusses the main features of the estimation and testing procedures implemented in the paper. Section 4 presents the main results. Finally, Section 5 summarizes and concludes.

2 Data

Our dataset consists of four market-wide climate risk factors sampled daily over the period from January 3, 2000, to January 31, 2025.² These factors are constructed using LDA machine learning applied to a broad corpus of news articles available from the Thomson Reuters News Archive for Machine Readable News. In addition, we consider daily VIX data obtained from the FRED database of the Federal Reserve Bank of St. Louis over the same period.

LDA is an unsupervised machine learning modeling technique designed to uncover latent categories (“topics”) in a large collection of documents under the assumption that each of these documents is a mixture of several topics that can be characterized by certain word probabilistic distributions. The goal of LDA is to identify these topics and determine the distribution of topics within each document and the distribution of words within each topic. To this end, LDA decomposes the corpus of interest into a pre-specific number of topics on the basis of the distribution of the unique words (the “vocabulary”) appearing in the text. In contrast to supervised methods, in which the researcher specifies the particular categories *ex ante*, LDA only requires the number of topics to be set in

²The data are provided by R. Faccini and G. Skiadopoulos. Sources: R. Faccini (<https://sites.google.com/site/econrenatofaccini>) and G. Skiadopoulos (<https://sites.google.com/view/george-skiadopoulos>).

advance. The latent topics are inferred on a probabilistic basis without prior labeling using Bayesian principles. The resulting categories can subsequently be labeled according to the unique words that occur most frequently within each topic. Additionally, LDA represents each document as a probability-weighted mixture of topics, where each weight measures the relative frequency with which a given topic appears in that document. Summing these topic shares across all articles published on a given day yields a daily measure of the intensity of news coverage for each of the topics identified by the LDA algorithm. This is the methodological basis underlying the construction of climate risk factors; see [Faccini et al. \(2023\)](#) and references therein for further details.

The textual analysis conducted in [Faccini et al. \(2023\)](#) is restricted to the universe of Reuters articles written in English in which the terms *climate change* or *global warming* appear at least once. Given this corpus, LDA is implemented with a number of 25 topics. Among these resultant categories, the authors select four topics that admit a clear interpretation as market-wide climate risk factors associated with either physical or transition risk and that are relevant for investors in U.S. equities. These categories are briefly described below (see [Faccini et al. \(2023\)](#) for greater details):

1. U.S. Climate Policy (Transition Risk). This factor captures news about the U.S. political debate and legislative actions related to climate change, including elections, proposed or passed bills, regulatory initiatives, and key political appointments. It represents imminent transition risk, as U.S. policy decisions can affect firms' costs and profitability over relatively short horizons.
2. International Climate Summits (Transition Risk). This factor reflects news about international negotiations and agreements on climate policy (e.g., the Kyoto Protocol, the Copenhagen Summit, and UN climate conferences). It proxies transition risks arising from the possibility of coordinated global policy actions, which tend to materialize slowly due to negotiation and implementation lags.
3. Natural Disasters (Physical Risk). This factor captures news linking climate change to extreme weather events such as floods, droughts, wildfires, hurricanes, and storms. It reflects investors' concerns about the increasing frequency and severity of climate-related disasters worldwide, interpreted as long-term physical risks associated with climate change.
4. Global Warming (Physical Risk). This factor measures news coverage related to rising global temperatures and greenhouse gas emissions, including scientific reports and assessments by organizations such as the World Meteorological Organization. It captures gradual, long-horizon physical risks stemming from climate change itself rather than discrete events.

As discussed in [Faccini et al. \(2023\)](#), these factors are weakly correlated at the daily frequency and capture different dimensions of climate risk. Interestingly, only the U.S. climate policy factor is found to be priced cross-sectionally in U.S. stock returns, suggesting that investors primarily price transition risk from government intervention, rather than long-term physical climate risks. Figure 1 and Figure 2 show the dynamics of each of the LDA-based climate risk factors and their sample autocorrelation functions, respectively. These variables exhibit highly volatile dynamics subject to extreme movements, resulting in large kurtosis and positive skewness.³ Similar to high-frequency financial returns, daily climate risk factors display time-varying conditional heteroskedasticity, consistent with clustering in media attention. At the same time, the four series exhibit strong persistence, as their sample autocorrelations decay slowly toward zero, a distinctive feature of fractionally integrated processes.

[Insert Figures 1 and 2 around here]

Given that our main interest lies in the long-run component of these series, we conduct the analysis using the logarithm of monthly averages of the daily climate risk factors. This transformation smooths high-frequency transitory fluctuations and mitigates the influence of extreme observations and conditional heteroskedasticity, thereby facilitating more accurate inference on long-run dependence while remaining consistent with the economic interpretation of climate risk as a slowly evolving state variable.⁴ Table 1 reports descriptive statistics (Panel A) and sample correlations (Panel B) for the four climate risk series and the VIX. Within the climate series forming the $\mathcal{CR}$ group, the log-transformed monthly averages exhibit statistically significant positive correlations ranging from 0.361 (Natural disasters and International summits) to 0.640 (U.S. Climate policy and Natural disasters), suggesting the presence of common long-term components across these series. The climate risk series are also mildly positively correlated with the VIX. It should be noted, however, that correlation analysis is only formally valid under stationarity, an assumption that need not hold for the series under consideration. Accordingly, these correlations are interpreted with caution, and no strong conclusions are drawn from this evidence alone.

[Insert Table 1 around here: Descriptive statistics]

³Extreme observations resembling additive outliers are known to induce a downward bias in estimates of the fractional integration parameter; see, among others, [Tolvi \(2003\)](#) and [Haldrup and Nielsen \(2007\)](#).

⁴Although this approach entails a reduction in the total number of observations, removing high-frequency components that may be both distortive and uninformative is sensible ([Hualde and Nielsen, 2023](#)). It is widely acknowledged in the unit root testing literature that a long time span is more informative than a large number of high-frequency observations, since identification relies on low-frequency variation; see, among others, [Shiller and Perron \(1985\)](#) and [Campbell and Perron \(1991\)](#). The same consideration naturally extends to the identification of fractional unit roots.

3 Methodology

Long-range persistence in the sense of slowly decaying autocorrelations is a stylized fact in many time series from economics and finance (Hassler and Pohle, 2023). The class of fractionally integrated models is the most common approach for capturing this pattern of dependence in the time series analysis. A variable x_t is said to be a Type-II fractionally integrated process with parameter d , denoted $I(d)$, if $(1 - L)^d x_t = u_t I(t > 0)$, where L is the lag operator, $I(\cdot)$ is the indicator function, and u_t is a covariance stationary process. The fractional differencing operator $(1 - L)^d$ is defined as: $(1 - L)^d = \sum_{j=0}^{\infty} \omega_j(d) L^j$, where $\omega_j(d) = \Gamma(j - d) / [\Gamma(j + 1) \Gamma(-d)]$ and $\Gamma(\cdot)$ is the gamma function. This setting allows for fractional values of the d parameter, encompassing $I(1)$ (integrated/unit root, $d = 1$) and $I(0)$ (weakly dependent, $d = 0$) models as special cases. For non-integer values $0 < d < 1/2$, the series x_t exhibits (asymptotic) stationary long-range dependence characterized by autocorrelations that decay hyperbolically to zero, whereas values $d \geq 1/2$ characterize non-stationary persistence; see Hualde and Nielsen (2023) for a review. In a multivariate framework, long-range dependence is often driven by common persistent stochastic trends, which motivates the (fractional) cointegration analysis in the general terms discussed in Engle and Granger (1987). In the following subsections, we detail the estimation and testing procedures implemented in our analysis.

3.1 Univariate estimation

The fractionally integrated parameter d of a univariate series can be estimated in different ways. Semiparametric methods in the frequency domain are very popular because, unlike fully parametric approaches, they do not require the specification of the short-run dynamics of the process. Within this category, the Exact Local Whittle (ELW) estimator developed by Shimotsu and Phillips (2005) and Shimotsu (2010) achieves greater efficiency than the log-periodogram regression estimator of Geweke and Porter-Hudak (1983). Furthermore, unlike the log-periodogram regression estimator and the local Whittle estimator of Kuensch (1987), the ELW is consistent and asymptotically normally distributed for any value of d . The ELW estimator of d is determined as:

$$\hat{d}_{ELW} = \arg \min_{d \in \mathcal{D}} \left[\log \overline{G(d)} - 2d \frac{1}{m} \sum_{j=1}^m \log \lambda_j \right] \quad (1)$$

where the bandwidth $m = m(T)$ determines the number of low-frequency periodogram ordinates used in the estimation, $\mathcal{D}$ is a bounded admissible parameter space of d , and:

$$\overline{G(d)} = \frac{1}{m} \sum_{j=1}^m I_{(1-L)^d x}(\lambda_j) \lambda_j^{2d}, \quad (2)$$

where:

$$I_{(1-L)^d x}(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T (1-L)^d x_t e^{i\lambda_j t} \right|^2 \quad (3)$$

is the periodogram of $(1-L)^d x_t$ evaluated at the fundamental frequency $\lambda_j = 2\pi j/T$, with $|\cdot|$ denoting the modulus of the argument.

Under the set of assumption in Shimotsu and Phillips (2005), $\hat{d}_{ELW}$ is consistent and asymptotically normal distributed, $m^{1/2}(\hat{d}_{ELW} - d_0) \rightarrow_d \mathcal{N}(0, 1/4)$, with d_0 denoting the true value. Consistency is achieved by allowing the bandwidth m to slowly diverge with the sample size T . Any $m = [T^\alpha]$, with $0 < \alpha < 1$ and $[\cdot]$ denoting the integer part of the argument, verifies the conditions in Shimotsu and Phillips (2005). In our analysis, we set $\alpha \in \{0.60, 0.65\}$.

Shimotsu (2010) extends this setting to deal with the presence of deterministic components considering $x_t = \delta(t) + x_t^*$, where $(1-L)^d x_t^* = u_t$ and $\delta(t) = \sum_{i=0}^s \delta_i t^i$ is a polynomial time trend. To avoid discontinuities in the asymptotic theory in the relevant cases $d = 0$ and $d = 1$, Shimotsu (2010) proposes a two-stage feasible ELW estimation (FELW2S) building on the OLS residuals $\hat{x}_t^* = x_t - \hat{\delta}(t)$ from a detrending regression. In the first stage, the procedure considers a preliminary $\sqrt{m}$ -consistent estimate $\tilde{d}$ given $\hat{x}_t^*$, such as the tapered Whittle estimator of Velasco (1999). This estimate is consistent and asymptotically Gaussian, but it has greater variance than the final estimator. In the second stage, the FELW2S results from optimizing the feasible ELW objective function (i.e., considering $\hat{x}_t^*$ instead of x_t^* in (1)) in a neighborhood of $\tilde{d}$. This can be obtained from a finite number of iterations in the Newton-Raphson algorithm. This allows the resultant estimator to achieve the same variance as the original ELW estimator when there is no deterministic component. More specifically, as T is allowed to diverge, $m^{1/2}(\hat{d}_{FELW2S} - d_0) \xrightarrow{d} \mathcal{N}(0, 1/4)$ for the process with $-0.5 < d_0 < 2$ and no polynomial trend, and for the process with $-0.5 < d_0 < 1.75$ and polynomial trend. Because of the greater generality of this procedure, we implement this estimation in our analysis, dealing with the possibility of a constant mean ($\delta(t) = \delta_0$) and a linear time trend ($\delta(t) = \delta_0 + \delta_1 t$).

3.2 Fractional cointegration

A p -vector time series $\mathbf{X}_t$ is said to be fractionally cointegrated $CI(d, b)$ if each element of $\mathbf{X}_t$ is $I(d)$ and there exists a system of r linearly independent combinations, $\mathbf{Z}_t = \beta' \mathbf{X}_t$ with $\beta \neq 0$, such that each element of $\mathbf{Z}_t$ (denominated cointegrating error) is $I(d - b)$, where $b > 0$ denotes the order of reduction obtained by the linear combinations. The r rows of the matrix β' are the cointegrating vectors, and the number of such relations, $1 \leq r < p$, is called the cointegrating rank. A system of p variables with rank $r \geq 1$ is driven by $p - r$ common $I(d)$ stochastic trends and possesses r long-run equilibrium relations $\mathbf{Z}_t = 0$, each exhibiting $I(d - b)$ behavior.

The existence of cointegration is formally addressed by determining the cointegrating rank of the system. In our analysis, we implement two complementary methods. First, we utilize the semiparametric procedure in [Nielsen and Shimotsu \(2007\)](#) to test the necessary condition that all fractional integration coefficients are equal and to estimate the cointegrating rank robustly. Subsections 3.2.1 and 3.2.2 describe this procedure. In addition, we estimate the FCVAR model of [Johansen and Nielsen \(2010\)](#) by Maximum Likelihood. The FCVAR procedure is described in Section 3.2.3.

3.2.1 Testing coefficient homogeneity

A crucial requirement for a system $\mathbf{X}_t$ to be fractionally cointegrated as defined in [Engle and Granger \(1987\)](#) is that all variables must share the same fractional integration order.⁵ A significant difficulty in this context is that the asymptotic behavior of joint tests involving fractional coefficients is affected by the presence or absence of fractional cointegration. Since it is formally unknown whether the variables in the system are cointegrated, inference must be consistent against this consideration. Under the assumption $|d| < 1/2$, [Robinson and Yajima \(2002\)](#) provided a robust test for coefficient homogeneity in a possibly cointegrated system and discussed a method to consistently estimate the cointegrating rank. [Nielsen and Shimotsu \(2007\)](#) extended this setting to accommodate both (asymptotic) stationary and nonstationary fractionally integrated processes, which is discussed in detail below.

Let $\mathbf{d} = (d_1, \dots, d_p)'$ denote the vector of fractionally integrated parameters. The hypothesis of homogeneity of fractional coefficients can be stated as $H_0 : \mathbf{S}\mathbf{d} = 0$, where $\mathbf{S} = (\mathbf{I}_{p-1}, -\mathbf{1}_{p-1})$, with $\mathbf{I}_s$ denoting the s -identity matrix, and $\mathbf{1}_s$ a s -vector of ones. To test this hypothesis, [Nielsen and Shimotsu \(2007\)](#) propose the following Wald-type test involving $p - 1$ restrictions:

$$\Lambda_T = m(\mathbf{S}\hat{\mathbf{d}})' \left[\frac{1}{4} \mathbf{S} \hat{\mathbf{D}}^{-1} (\hat{\mathbf{G}} \circ \hat{\mathbf{G}}) \hat{\mathbf{D}}^{-1} \mathbf{S}' + h^2 \mathbf{I}_{p-1} \right]^{-1} (\mathbf{S}\hat{\mathbf{d}}),$$

where $\circ$ denotes the Hadamard product, $\hat{\mathbf{d}}$ is a $\sqrt{m}$ -consistent estimate of $\mathbf{d}$ using a bandwidth parameter m ; $\hat{\mathbf{G}}$ is a consistent estimator of $\mathbf{G}$ (the limit of the spectral density matrix of the d -th differenced variables as $\lambda \rightarrow 0^+$), $\hat{\mathbf{D}} = \text{diag}(\hat{\mathbf{G}})$ is a diagonal matrix, and $h = (\log T)^{-a}$ for some chosen $a > 0$. The test exploits $m^{1/2}(\hat{\mathbf{d}} - \mathbf{d}_0) \rightarrow_d \mathcal{N}(0, \mathbf{V})$, with $\mathbf{V} = \frac{1}{4} \mathbf{D}^{-1} \mathbf{G} \circ \mathbf{G} \mathbf{D}^{-1}$, and $\mathbf{D} = \text{diag}(\mathbf{G})$. As discussed by [Nielsen and Shimotsu \(2007\)](#), the elements of $\mathbf{d}$ can be estimated consistently using univariate (two-step feasible) ELW, while $\hat{\mathbf{G}} = \sum_{j=1}^m \text{Re} [I_{\Delta \hat{\mathbf{d}} \mathbf{X}}(\lambda_j)] / m$ is a consistent estimate of $\hat{\mathbf{G}}$, where

⁵A more general definition of fractional cointegration allows different orders of integration for each individual series; see [Hualde and Nielsen \(2023\)](#).

$\text{Re} [I_{\Delta \hat{\mathbf{d}} \mathbf{X}}(\lambda_j)]$ denotes the real part of the coperiodogram of $(1 - L)^{\hat{\mathbf{d}}} \mathbf{X}_t$ evaluated at the frequency λ_j .

If there is no cointegration, $\mathbf{G}$ is of full rank and non-singular. Because the sequence $h = h(T)$ becomes asymptotically negligible, Λ_T converges weakly to a χ_{p-1}^2 distribution as $T \rightarrow \infty$ owing to the asymptotic normality of $\hat{\mathbf{d}}$. Under cointegration, $\mathbf{G}$ is singular and Λ_T converges in probability to zero, with the regularization term h ensuring this convergence despite the reduced-rank nature of $\mathbf{G}$; see [Nielsen and Shimotsu \(2007\)](#). As a result, a significantly large value of Λ_T with respect to a χ_{p-1}^2 can be taken as an evidence against the equality of the integration orders, irrespectively of whether the variables are fractionally cointegrated or not. Monte Carlo analysis reported in [Nielsen and Shimotsu \(2007\)](#) shows that $h = (\log T)^{-1/2}$ outperforms other alternatives, so we implement this selection in our analysis.

3.2.2 Semiparametric estimation of the cointegrating rank

The rank of the spectral density matrix plays a fundamental role in the determination of the cointegrating rank. If $\text{rank}(\mathbf{G}) = p$, then $r = 0$ and there is no cointegration even though all the elements of X_t are $I(d)$. On the other hand, $\text{rank}(\mathbf{G}) = p - r$ with $1 \leq r < p$ if and only if there are r cointegrating relations ([Granger, 1986](#)). Since the rank of a matrix is equal to the number of its non-zero eigenvalues, the analysis of the eigenvalues of a consistent estimation of $\mathbf{G}$ under H_0 allows us to consistently estimate r .

Building on this estimation strategy, [Nielsen and Shimotsu \(2007\)](#) propose a model selection procedure to estimate r by estimating $\mathbf{G}$ and its eigenvalues. Let $\bar{d}$ be the average of the univariate $\sqrt{m}$ -consistent estimates $\hat{d}_1, \dots, \hat{d}_p$. A suitable consistent estimate of $\mathbf{G}$ for inference in this context, say $\widehat{\mathbf{G}}(\bar{d})$, can then be obtained by averaging m_1 multivariate periodogram ordinates of the p -vector process $(1 - L)^{\bar{d}} X_t$, with m_1 chosen such that $m_1/m \rightarrow 0$. Then, as suggested in [Robinson and Yajima \(2002\)](#), the cointegrating rank can be estimated as:

$$\hat{r} = \arg \min_{u=0, \dots, p-1} L(u); \quad L(u) = v(T)(u - p) - \sum_{i=1}^{p-u} \hat{\delta}_i, \quad (4)$$

where $\hat{\delta}_i$ denotes the i -th eigenvalue (sorted in descending order) of $\widehat{\mathbf{G}}(\bar{d})$, and some $v(T) > 0$ satisfying $v(T) + m_1^{-1/2} v(T)^{-1} \rightarrow 0$ as $T \rightarrow \infty$. Any sequence of the form $v(T) = m_1^{-\kappa}$ with $0 < \kappa < 1/2$ verifies this restriction. As noticed in [Nielsen and Shimotsu \(2007\)](#), in a finite sample larger values of $v(T)$ tend to yield larger estimates $\hat{r}$, whereas small values $v(T)$ lead to excessively low estimates r . Unfortunately, the estimation is fairly sensitive to this choice and the authors recommend using different choices of $v(T)$. In our analysis, we set $m_1 = \lceil T^{0.5} \rceil$ and consider values $v(T) = m_1^{-\kappa}$ with $\kappa \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$.

3.2.3 Fractionally Cointegrated VAR (FCVAR) model

To introduce the FCVAR model, it is useful to first consider the non-fractional case $I(1)/I(0)$ characterized by $d = b = 1$. In this setting, the Granger Representation Theorem (Engle and Granger, 1987) states that a cointegrated system admits a Vector Error Correction Model (VECM) representation: a proportion of the disequilibrium (i.e., the cointegrating error) from one period is corrected in the next period. Thus, if $\mathbf{Y}_t$ is a p -dimensional $I(1)$ time series with cointegrating rank r , the Cointegrated VAR (CVAR) model is characterized by the following error correcting representation:

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{Y}_{t-1} + \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta \mathbf{Y}_{t-i} + \boldsymbol{\epsilon}_t \quad (5)$$

where $\Delta = (1 - L)$ is the difference operator, and it is assumed that $\boldsymbol{\epsilon}_t$ is p -dimensional independent and identically distributed with mean zero and covariance matrix $\boldsymbol{\Omega}$. The unknown terms $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(p \times r)$ matrices. The former contains the adjustment factors (long-run loadings) capturing how error correction takes place, while the latter contains the cointegrating vectors of the system. The $(p \times p)$ $\boldsymbol{\Gamma}_i$ matrices capture stationary short-run dynamics through multivariate VAR(k) augmentation.

Noting that $\mathbf{Y}_{t-1} = L \mathbf{Y}_t$ and that $L = 1 - \Delta$, Johansen and Nielsen (2012) derive the FCVAR model by replacing the difference and lag operators (Δ and L) in the previous expression by their fractional counterparts, $\Delta^b = (1 - L)^b$ and $L_b = 1 - \Delta^b$, respectively, obtaining:

$$\Delta^b \mathbf{Y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' L_b \mathbf{Y}_t + \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta^b L_b^i \mathbf{Y}_t + \boldsymbol{\epsilon}_t \quad (6)$$

This model is applied to the $CI(d, b)$ system by setting $\mathbf{Y}_t = \Delta^{d-b} \mathbf{X}_t$. Substituting this identity into the equation above yields the FCVAR model expressed in terms of the original $I(d)$ series, $\mathbf{X}_t$, namely:

$$\Delta^d \mathbf{X}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' L_b \Delta^{d-b} \mathbf{X}_t + \sum_{i=1}^k \boldsymbol{\Gamma}_i L_b^i \Delta^d \mathbf{X}_t + \boldsymbol{\epsilon}_t \quad (7)$$

In this characterization, the main parameters of interest $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ have the same interpretations from the CVAR model. The coefficients in $\boldsymbol{\alpha}$ represent the speed of adjustment towards equilibrium for each of the variables, while the rows of $\boldsymbol{\beta}'$ are the cointegrating relationships in the system, and $r = \text{rank}(\boldsymbol{\alpha} \boldsymbol{\beta}')$ is the cointegrating rank. The unknown parameters in the FCVAR model $(d, b, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_k)$ are estimated by the maximum likelihood-based procedures described in Johansen and Nielsen (2012) and Nielsen and Morin (2014). In the empirical implementation of this model, we allow for

a deterministic drift $\boldsymbol{\mu}$, replacing $\mathbf{X}_t$ in equation (7) with $\mathbf{X}_t^* = \mathbf{X}_t - \boldsymbol{\mu}$ and estimating this parameter vector jointly with the remaining parameters.

4 Empirical results

4.1 Long-memory estimation

Table 2 reports univariate FELW2S estimates of the fractional integration parameter, $\hat{d}$, for the four climate risk factors and the VIX. The table also presents standard errors, 95% confidence intervals, and t -tests for the null hypotheses $H_0 : d = 0$ and $H_0 : d = 1$. Estimation is carried out using monthly time series constructed as the natural logarithm of the arithmetic mean of the daily observations within each calendar month. In implementing the estimation procedure, we consider two alternative specifications for the deterministic component: a constant only (Panel A), and a constant and linear time trend (Panel B). In both cases, bandwidths are selected as $m = [T^\alpha]$, with $\alpha \in \{0.60, 0.65\}$. The bottom of each panel reports the Wald-type test statistic Λ_T for coefficient homogeneity, discussed in Subsection 3.2.1, applied to the series in the $\mathcal{CR}$ system (climate risk factors only) and the $\mathcal{VCR}$ system (climate risk factors and the VIX).

[Insert Table 2 around here: Long-memory parameters]

FELW2S estimation provides strong evidence in favor of fractional integration in the four climate risk measures as well as in the VIX. The results are robust across alternative specifications of the deterministic component and different bandwidth choices. For each climate risk factor, the null hypotheses $H_0 : d = 0$ (i.e., weak dependence) and $H_0 : d = 1$ (i.e., an autoregressive unit root process) are strongly rejected. Although point estimates of d tend to exceed the 0.5 threshold in most cases, the 95% confidence intervals typically span both the stationary and non-stationary regions of the parameter space, preventing definitive conclusions regarding the stationarity of the underlying series. This outcome is well known in the empirical analysis of long-memory financial variables; see, for instance, [Balboa et al. \(2021\)](#). Consistent with previous findings in the literature, a qualitatively similar pattern emerges from the univariate analysis of the long-run dynamics of the VIX.

Overall, the evidence suggests that both the $\mathcal{CR}$ and the $\mathcal{VCR}$ systems are likely driven by non-stationary fractionally integrated dynamics characterized by a common fractional exponent. While there is no prior evidence for climate risk factors, the presence of non-stationary long-memory dynamics in the VIX over extended time horizons is consistent with the existing literature; see, among others, [Bandi and Perron \(2006\)](#), [Hassler et al. \(2016\)](#), and [Balboa et al. \(2021\)](#). Given the similarity of the parameter estimates, it is not surprising that the hypothesis that the four textual climate risk factors and VIX are driven by a long-memory model with a common fractional exponent cannot be rejected.

This evidence holds across both specifications of the deterministic component and the two bandwidth choices. Notably, the p -values of the Wald-type test statistic Λ_T are particularly large for $m = \lceil T^{0.65} \rceil$.

4.2 Fractional cointegrating rank

The previous findings motivate further interest in the possibility of fractional cointegration in both the $\mathcal{CR}$ and $\mathcal{VCR}$ systems. We formally address this issue in this section by estimating the cointegrating rank using a semiparametric approach and estimating the vector error-correction model representation.

4.2.1 Semiparametric estimation

We first report and discuss the estimates of the cointegrating rank, $\hat{r}$, obtained using the model selection procedures proposed by Robinson and Yajima (2002) and Nielsen and Shimotsu (2007); see Subsection 3.2.2. To save space, we report results only for the specification that includes a linear time trend (results for the model with a constant are qualitatively similar and available upon request).

We rely on the univariate FELW2S estimates (based on bandwidth choices $m = \lceil T^\alpha \rceil$ with $\alpha \in \{0.60, 0.65\}$) and on the spectral density matrix of the $\bar{d}$ -th differenced process (with $\bar{d}$ given by the arithmetic mean of the univariate estimates) around the origin computed using a bandwidth $m_1 = \lceil T^{0.5} \rceil$. In the model selection procedure, we follow Nielsen and Shimotsu (2007) and compute the eigenvalues of the correlation matrix $\mathbf{P} = \widehat{\mathbf{D}}^{-1/2} \widehat{\mathbf{G}} \widehat{\mathbf{D}}^{-1/2}$, setting the threshold $v(T) = m_1^{-\kappa}$ with $\kappa \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$ in order to address the sensitivity of results to this choice. For each of the (m, κ) combinations, Panels A and B of Table 3 report the estimated cointegrating rank together with the values $L(u)$, $u = 0, \dots, p-1$ for variables in the $\mathcal{CR}$ system ($p = 4$) and the VIX-augmented $\mathcal{VCR}$ system ($p = 5$), respectively.

[Insert Table 3 around here: NS cointegrating rank]

The results indicate the presence of at least one fractional cointegrating relationship in both the $\mathcal{CR}$ and $\mathcal{VCR}$ systems. This evidence supports the view that, although policy-related and scientific narratives may dominate specific dimensions of climate discourse, there are common underlying stochastic trends in the textual climate risk factors. Furthermore, the existence of fractional cointegration in the VIX-augmented system supports the hypothesis that climate-risk factors comove in the long-term with VIX and, hence, are partially driven by market sentiment. This evidence provides support from a different perspective to the results in Faccini et al. (2023), justifying why climate-risk is priced.

While this semiparametric approach can infer the cointegrating rank without estimating the cointegrating vectors, it requires the choice of bandwidth and thresholds. Unfortunately, and as in the empirical analysis in [Nielsen and Shimotsu \(2007\)](#), the estimation results in our analysis prove sensitive to $v(T)$, which prevents us from drawing a conclusive picture on the number of cointegrating relations. We observe the same finite-sample pattern discussed in [Nielsen and Shimotsu \(2007\)](#), with smaller values of $v(T)$ leading to larger values of $\hat{r}$. According to the semiparametric model selection procedure, the cointegrating rank may be as high as three in both the $\mathcal{CR}$ and $\mathcal{VCR}$ systems.

4.2.2 FCVAR estimation

In order to shed further light on the number of cointegrating relations in the $\mathcal{VCR}$ system, we also implement the testing procedure of [Johansen and Nielsen \(2012\)](#) and [Nielsen and Morin \(2014\)](#) in the FCVAR setting relying on maximum-likelihood (ML) estimation. More specifically, we can test hypotheses of the form $H_0 : \text{rank}(\mathbf{\Pi}) = r$ against the alternative $H_1 : \text{rank}(\mathbf{\Pi}) = p$, with $\mathbf{\Pi} = \mathbf{\alpha}\mathbf{\beta}'$, using likelihood ratio (LR) tests by maximizing the profile likelihood function under both hypotheses. The asymptotic distribution of the LR test depends both qualitatively and quantitatively on the cointegrating gap b and on the nuisance term $q = p - r$.

[Insert Table 4 around here: FCVAR rank]

In Table 4 we show the main results of the cointegration rank test by [Johansen and Nielsen \(2012\)](#) in the FCVAR setting given a lag-order $k = 1$. The null hypothesis of no cointegration ($r = 0$) leads to a LR test of 61.54 and, therefore, is strongly rejected. The null hypothesis that the cointegrating rank is equal to one ($r = 1$) is marginally rejected at the 10% significance level. On the other hand, when testing the null hypothesis that $r = 2$, the LR statistic is significantly smaller than the corresponding critical value as reflected in p -value equal to 0.50. Consequently, whereas the statistical evidence supporting a single cointegration relation is somewhat borderline, the LR testing approach favors the existence of two long-term equilibrium relations in the $\mathcal{VCR}$ system.

Given $r=2$, the estimates of the long-memory parameter and the cointegrating gap are $\hat{d} = 0.773$ and $\hat{b} = 0.207$, respectively. These values do not seem to differ from those observed for $r=1$. The value of the long-memory coefficient, slightly greater than the average values estimated by univariate FEWL2S for climate-risk series but remarkably close to the values of the VIX, shows nonstationary dynamics. The value of the cointegrating gap fall in the region $0 < b < 1/2$ corresponding to the so-called ‘weak cointegration’. Note that, for values of b in this region, the asymptotic distribution of the LR test is a Chi-squared distribution with q^2 degrees of freedom.

5 Concluding remarks

This paper investigates the stochastic properties and long-run equilibrium relationships of textual climate risk factors derived from news-based measures. Using four climate risk indices constructed from Reuters news via Latent Dirichlet Allocation as proposed in [Faccini et al. \(2023\)](#), we document strong evidence supporting fractional integration building on ELW estimation ([Shimotsu and Phillips, 2005](#); [Shimotsu, 2010](#)). The size of the fractional exponent in the climate-risk factors analyzed is similar and comparable to that observed in market-wide implied volatility. Hence, the long-term underlying dynamics in climate-related narratives in the media seem to be driven by persistent, non-stationary clustering effects similar to those that characterize time-varying volatility of financial returns in stock markets.

We further analyze the multivariate structure of climate risk factors using both semiparametric ([Robinson and Yajima, 2002](#); [Nielsen and Shimotsu, 2007](#)) and likelihood-based approaches ([Johansen and Nielsen, 2012](#); [Nielsen and Morin, 2014](#)) intended to detect fractional cointegration as generally defined in [Engle and Granger \(1987\)](#). Semiparametric methods indicate the presence of common stochastic trends, although the precise number of cointegrating relations is sensitive to threshold choices, a feature already documented in the previous literature. To provide a sharper characterization of the long-run structure, we estimate a fractionally cointegrated VAR model. The FCVAR results point to up to two non-stationary fractional cointegrating relations in a weakly cointegrating system. Overall, the existence of fractional cointegration the VIX-augmented system supports the interpretation that textual climate risk factors are persistent processes (partially) driven by market-wide sentiment. This finding complements existing asset-pricing results and helps understanding why certain dimensions of climate risk are priced in financial markets.

The existence of nonstationary fractional integration dynamics in climate risk factors is a relevant issue with a number of implications for the econometric analysis. For instance, the OLS regression of returns (which are well-known cases of $I(0)$ series) on a non-stationary $I(d)$ climate-risk factor can be framed into an unbalanced regression problem. The analysis of the consequences on asset pricing tests seems an interesting avenue left for future research.

References

- Afzal, A. and Sibbertsen, P. (2021). Modeling fractional cointegration between high and low stock prices in Asian countries. *Empirical Economics*, 60(2):661–682.
- Allahdadi, M. R., Fretheim, T., and Vindedal, K. (2024). Value of climate change news: A textual analysis. *Global Finance Journal*, 63:101052.
- Ardia, D., Bluteau, K., Boudt, K., and Inghelbrecht, K. (2023). Climate change concerns and the performance of green vs. brown stocks. *Management Science*, 69(12):7607–7632.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59.
- Balboa, M., Rodrigues, P. M., Rubia, A., and Taylor, A. R. (2021). Multivariate fractional integration tests allowing for conditional heteroskedasticity with an application to return volatility and trading volume. *Journal of Applied Econometrics*, 36(5):544–565.
- Bandi, F. M. and Perron, B. (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, 4(4):636–670.
- Bekaert, G., Hoerova, M., and Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788.
- Berger, D., Chaboud, A., and Hjalmarsson, E. (2009). What drives volatility persistence in the foreign exchange market? *Journal of Financial Economics*, 94(2):192–213.
- Bua, G., Kapp, D., Ramella, F., and Rognone, L. (2024). Transition versus physical climate risk pricing in European financial markets: A text-based approach. *The European Journal of Finance*, 30(17):2076–2110.
- Campbell, J. Y. and Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. In Blanchard, O. J. and Fischer, S., editors, *NBER Macroeconomics Annual 1991*, pages 141–201. MIT Press, Cambridge, MA.
- Caporin, M., Ranaldo, A., and De Magistris, P. S. (2013). On the predictability of stock prices: A case for high and low prices. *Journal of Banking & Finance*, 37(12):5132–5146.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.
- Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251–276.

- Faccini, R., Matin, R., and Skiadopoulos, G. (2023). Dissecting climate risks: Are they reflected in stock prices? *Journal of Banking & Finance*, 155:106948.
- Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.
- Giglio, S., Kelly, B., and Stroebe, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13(1):15–36.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227–238.
- Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics & Statistics*, 48(3):213–228.
- Haldrup, N. and Nielsen, M. Ø. (2007). Estimation of fractional integration in the presence of data noise. *Computational Statistics & Data Analysis*, 51(6):3100–3114.
- Hassler, U. and Pohle, M.-O. (2023). Forecasting under long memory. *Journal of Financial Econometrics*, 21(3):742–778.
- Hassler, U., Rodrigues, P. M., and Rubia, A. (2016). Quantile regression for long memory testing: a case of realized volatility. *Journal of Financial Econometrics*, 14(4):693–724.
- Hualde, J. and Nielsen, M. Ø. (2023). Fractional integration and cointegration. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.
- Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, 158(1):51–66.
- Johansen, S. and Nielsen, M. Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, 80(6):2667–2732.
- Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3):1067–1111.
- Kuensch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press) Utrecht.
- Nielsen, M. Ø. and Morin, L. (2014). Fcvarmodel.m: A Matlab software package for estimation and testing in the fractionally cointegrated VAR model. *Journal, (QED working paper 1273, Queen’s University)*.

- Nielsen, M. Ø. and Shimotsu, K. (2007). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. *Journal of Econometrics*, 141(2):574–596.
- Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106(2):217–241.
- Shiller, R. J. and Perron, P. (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters*, 18(4):381–386.
- Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2):501–540.
- Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, 33(4):1890–1933.
- Tolvi, J. (2003). Long memory and outliers in stock market returns. *Applied Financial Economics*, 13(7):495–502.
- Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.
- Wang, C.-W., Wu, Y.-C., Hsieh, H.-Y., Huang, P.-H., and Lin, M.-C. (2022). Does green bond issuance have an impact on climate risk concerns? *Energy Economics*, 111:106066.

Figures and Tables

Figure 1. Daily textual climate risk factors.

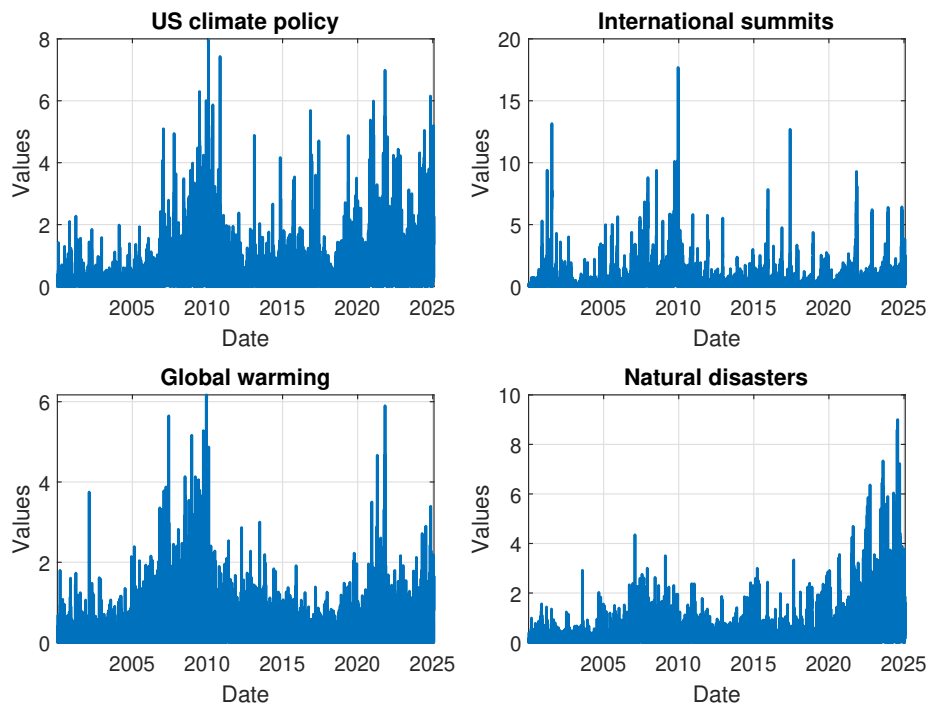


Figure 2. Sample autocorrelation function of daily textual climate risk factors.

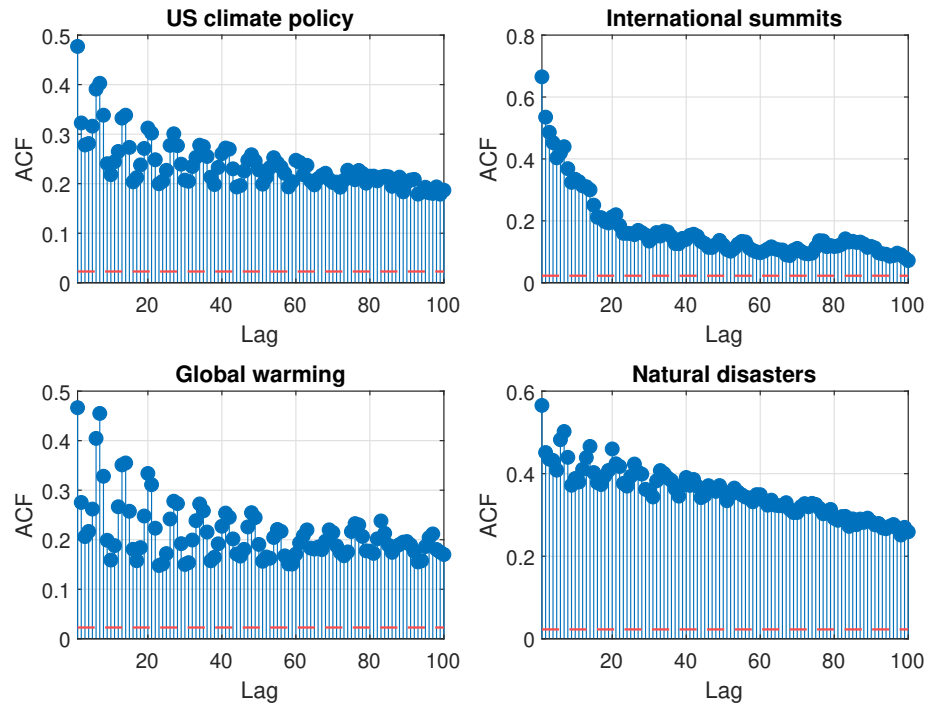


Table 1. Descriptive statistics (Panel A, by rows: mean, median, standard deviation, minimum, maximum, skewness, kurtosis, Jarque-Bera test for normality, Jarque-Bera p -value, and number of observations) and correlation matrix (Panel B) for the textual climate risk factors U.S. Climate Policy (USClimPol), Internal sumnits (IntSum), Global warming (GlobWarm), and the VIX index. All monthly variables in logs.

Panel A: Descriptive statistics					
	USClimPol	IntSum	GlobWarm	NatDis	VIX
Mean	-0.966	-1.135	-1.092	-1.250	2.924
Median	-0.941	-1.199	-1.110	-1.331	2.878
StdDev	0.874	0.951	0.664	0.903	0.347
Min	-3.015	-3.235	-2.807	-4.036	2.315
Max	0.950	1.643	0.562	1.195	4.138
Skewness	-0.253	0.227	0.101	0.237	0.685
Kurtosis	2.342	2.664	2.689	3.034	3.449
JB	8.645	3.994	1.724	2.842	26.068
p -value	0.013	0.136	0.422	0.241	0.000
Observations	301	301	301	301	301

Panel B: Correlation matrix					
	USClimPol	IntSum	GlobWarm	NatDis	VIX
USClimPol	1.000	0.506	0.583	0.640	0.103
IntSum	0.506	1.000	0.627	0.361	0.107
GlobWarm	0.583	0.627	1.000	0.454	0.168
NatDis	0.640	0.361	0.454	1.000	0.016
VIX	0.103	0.107	0.168	0.016	1.000

Table 2. Semiparametric frequency-domain estimation of the fractionally integrated parameter d and inference. Estimation accounts for a deterministic mean (Panel A) and a linear time trend (Panel B) using 2-stage FELW estimation. In both cases, we consider bandwidths $m = \lceil T^{0.60} \rceil$ and $m = \lceil T^{0.65} \rceil$. The table reports parameter estimates ($\hat{d}$), standard errors (se), 95% confidence intervals (95% CI), and t -tests for $H_0 : d = 1$ and $H_0 : d = 0$ (p -values in parenthesis). The bottom of each panel reports the Wald-type test Λ_T (p -values in parenthesis) for homogeneity of the fractional parameter in the series of the $\mathcal{CR}$ system ($\Lambda_{\mathcal{CR}}$) and $\mathcal{VCR}$ system ($\Lambda_{\mathcal{VCR}}$).

	Panel A: Constant						Panel B: Linear time trend					
	$m = \lceil T^{0.60} \rceil$						$m = \lceil T^{0.60} \rceil$					
	$\hat{d}$	se	95% CI	$H_0 : d = 1$	$H_0 : d = 0$		$\hat{d}$	se	95% CI	$H_0 : d = 1$	$H_0 : d = 0$	
USClimPol	0.723	0.091	[0.544,0.902]	-3.04 (0.00)	7.92 (0.00)		0.672	0.091	[0.493,0.851]	-3.60 (0.00)	7.36 (0.00)	
IntSum	0.406	0.091	[0.227,0.585]	-6.51 (0.00)	4.45 (0.00)		0.405	0.091	[0.227,0.584]	-6.51 (0.00)	4.44 (0.00)	
GlobWarm	0.673	0.091	[0.494,0.852]	-3.58 (0.00)	7.37 (0.00)		0.673	0.091	[0.494,0.852]	-3.58 (0.00)	7.38 (0.00)	
NatDis	0.656	0.091	[0.477,0.835]	-3.76 (0.00)	7.19 (0.00)		0.536	0.091	[0.357,0.715]	-5.09 (0.00)	5.87 (0.00)	
VIX	0.696	0.091	[0.517,0.875]	-3.33 (0.00)	7.62 (0.00)		0.695	0.091	[0.516,0.874]	-3.34 (0.00)	7.61 (0.00)	
	$\Lambda_{\mathcal{CR}} : 5.308 (0.15)$						$\Lambda_{\mathcal{CR}} : 4.169 (0.24)$					
	$\Lambda_{\mathcal{VCR}} : 5.743 (0.22)$						$\Lambda_{\mathcal{VCR}} : 5.303 (0.26)$					
	$m = \lceil T^{0.65} \rceil$						$m = \lceil T^{0.65} \rceil$					
	$\hat{d}$	se	95% CI	$H_0 : d = 1$	$H_0 : d = 0$		$\hat{d}$	se	95% CI	$H_0 : d = 1$	$H_0 : d = 0$	
USClimPol	0.644	0.079	[0.489,0.799]	-4.51 (0.00)	8.14 (0.00)		0.547	0.079	[0.392,0.702]	-5.73 (0.00)	6.92 (0.00)	
IntSum	0.604	0.079	[0.449,0.759]	-5.01 (0.00)	7.64 (0.00)		0.607	0.079	[0.452,0.762]	-4.98 (0.00)	7.67 (0.00)	
GlobWarm	0.656	0.079	[0.501,0.811]	-4.35 (0.00)	8.30 (0.00)		0.655	0.079	[0.500,0.810]	-4.36 (0.00)	8.29 (0.00)	
NatDis	0.663	0.079	[0.509,0.818]	-4.26 (0.00)	8.39 (0.00)		0.609	0.079	[0.454,0.764]	-4.95 (0.00)	7.70 (0.00)	
VIX	0.639	0.079	[0.484,0.794]	-4.57 (0.00)	8.08 (0.00)		0.621	0.079	[0.466,0.776]	-4.79 (0.00)	7.86 (0.00)	
	$\Lambda_{\mathcal{CR}} : 0.260 (0.97)$						$\Lambda_{\mathcal{CR}} : 0.686 (0.88)$					
	$\Lambda_{\mathcal{VCR}} : 0.233 (0.99)$						$\Lambda_{\mathcal{VCR}} : 0.710 (0.95)$					

Table 3. Cointegrating rank estimation based on the model selection procedure in [Nielsen and Shimotsu \(2007\)](#). Given the variables of the $\mathcal{CR}$ (Panel A) and $\mathcal{VCR}$ (Panel B) system, the cointegrating rank is estimated as $\hat{r} = \arg \min_{u=0, \dots, p-1} L(u)$; $L(u) = v(T)(u - p) - \sum_{i=1}^{p-u} \hat{\delta}_i$.

Panel A: $\mathcal{CR}$ system ($p = 4$)										
κ	$m = [T^{0.60}]$					$m = [T^{0.65}]$				
	0.45	0.35	0.25	0.15	0.05	0.45	0.35	0.25	0.15	0.05
$L(0)$	-2.882	-2.516	-2.030	-1.385	-0.528	-2.882	-2.516	-2.030	-1.385	-0.528
$L(1)$	-2.905	-2.630	-2.266	-1.782	-1.139	-2.891	-2.616	-2.252	-1.768	-1.126
$L(2)$	-2.803	-2.620	-2.377	-2.054	-1.626	-2.804	-2.621	-2.377	-2.055	-1.627
$L(3)$	-2.551	-2.459	-2.338	-2.176	-1.962	-2.537	-2.445	-2.324	-2.163	-1.948
$\hat{r}$	1	1	2	3	3	1	2	2	3	3

Panel B: $\mathcal{VCR}$ system ($p = 5$)										
κ	$m = [T^{0.60}]$					$m = [T^{0.65}]$				
	0.45	0.35	0.25	0.15	0.05	0.45	0.35	0.25	0.15	0.05
$L(0)$	-3.603	-3.145	-2.538	-1.731	-0.660	-3.603	-3.145	-2.538	-1.731	-0.660
$L(1)$	-3.639	-3.273	-2.787	-2.142	-1.285	-3.629	-3.263	-2.777	-2.132	-1.275
$L(2)$	-3.541	-3.267	-2.902	-2.418	-1.776	-3.544	-3.270	-2.905	-2.421	-1.779
$L(3)$	-3.281	-3.098	-2.855	-2.533	-2.104	-3.277	-3.094	-2.851	-2.528	-2.100
$L(4)$	-2.541	-2.449	-2.328	-2.166	-1.952	-2.537	-2.446	-2.324	-2.163	-1.949
$\hat{r}$	1	1	2	3	3	1	2	2	3	3

Table 4. Rank determination in FCVAR system.

Rank	d	b	log-L.	LR stat.	p -value
0	0.528	0.636	-445.890	61.541	0.00
1	0.774	0.177	-426.847	23.455	0.10
2	0.773	0.207	-419.305	8.370	0.50
3	0.745	0.010	-415.972	1.704	0.79
4	0.779	0.010	-415.127	0.015	0.90
5	0.785	0.152	-415.120	—	—