Text-Based Financial Stress and Sovereign Risk in Brazil

Pedro Igor Oliveira[*, a]

Victor L. Stilita[†, a]

Ramon Lima Ribeiro[‡, a]

[a]CAEN/UFC

Abstract This paper develops three text-based indicators of financial stress for Brazil using Google Trends search data and financial news from major Brazilian outlets between 2008 and 2025. The first indicator is based on a curated dictionary of Portuguese-language search terms, the second uses a machine-learning selection procedure, and the third is built from 33,881 news articles. We assess their empirical relevance for Brazil's 5-year sovereign CDS spread using local projections, Granger-causality tests, and regime-switching analysis. The dictionary-based index shows the most stable association with subsequent CDS widening. The news-based measure also contains useful information, especially in the Granger exercises, while the machine-learning specification is less consistent in this application. The results suggest that search-based and news-based indicators can complement conventional tools for monitoring sovereign financial stress in Brazil.

Keywords: Financial Stress; Sovereign Credit Risk; Text Analysis; Google Trends

JEL codes: G01, G15, C58

1. Introduction

Financial stress and uncertainty are central to the analysis of macroeconomic fluctuations, financial instability, and the transmission of shocks across markets (Ellsberg, 1961; Bernanke, 1983; Dixit, 1992; Abel and Eberly, 1994; Bloom, 2007, 2009). Although these concepts are familiar in both theory and policy discussions, they are not directly observable. Their empirical measurement therefore depends on proxies that capture changes in perceptions, expectations, and market conditions.

A broad strand relies on market-based indicators such as spreads, volatility, and other price-based measures (Louzis et al., 2012; Bordo et al., 2000; Hakkio and Keeton, 2009; Ferreira and Mattos, 2018; Gaglianone and Areosa, 2016; Holló et al., 2012; Vdovychenko and Oros, 2014; Monin, 2017; Park and Mercado, 2014; Lefort et al., 2024). Another strand uses unstructured data,

[*]araujo.oliveirapg@gmail.com

[†]vstilita@alu.ufc.br

[‡]ramonribeiropop@alu.ufc.br

especially news content and textual sentiment, to recover information about uncertainty and stress that may not be fully reflected in market variables (Baker et al., 2016; Andres-Escayola et al., 2025a; Püttmann, 2018). More recently, search data have also been used to proxy for attention, expectations, and information demand (Castelnuovo and Tran, 2017; Pratap and Priyaranjan, 2023).

In Brazil, the main empirical contributions follow these same lines. Existing measures of uncertainty and financial stress use market data (Ferreira and Mattos, 2018; Gaglianone and Areosa, 2016), news data (Fernandes and Pereira, 2025), or combinations of both (Batista et al., 2024). These studies provide useful benchmarks, but they leave less room for a direct measure of how economic agents allocate attention to stress-related topics in real time.

We develop text-based financial stress measures using two informational environments. The first is search behavior, observed through Google Trends. The second is professional news coverage from major Brazilian outlets. The use of search data is motivated by the literature that interprets search intensity as a proxy for investor attention and information demand (Pereira et al., 2020; Ramos et al., 2017). The use of news data follows a parallel literature showing that media content helps shape expectations and provides information about macro-financial conditions (Baker et al., 2016). More broadly, both data sources are consistent with the idea that information processing is costly and selective, as emphasized by the theory of rational inattention (Sims, 2003; Maćkowiak et al., 2023).

The paper has three related but distinct goals. The first is to measure financial stress using text-based data. The second is to examine whether those indicators contain information about subsequent movements in Brazil's 5-year sovereign CDS spread, which we use as our benchmark for sovereign risk repricing. The third is to compare alternative ways of constructing such indicators. These three exercises are connected, but they are not identical. The first is a measurement problem, the second is a predictive-content exercise, and the third is a methodological comparison. For that reason, we interpret the empirical results as evidence on usefulness and relative performance rather than as a structural model of sovereign-risk determination.

Our empirical strategy reflects this distinction. First, we construct a dictionary-based search index using a curated set of Portuguese-language queries related to financial stress. Second, we estimate a machine-learning version of the index using LASSO to select search terms with predictive content for equity-market deterioration. Third, we build a news-based index from the text of 33,881 financial news articles published by G1, Folha de S.Paulo, and Valor Econômico.

We also report a combined index that aggregates the three measures as an exploratory benchmark.

The results can be summarized in three points. First, the dictionary-based search index is the specification most closely associated with subsequent CDS widening in this application. Second, the news-based index also contains predictive information, especially in the Granger-causality exercises. Third, the regime-switching model estimated on the dictionary index assigns higher stress probabilities to the main turbulence episodes in the sample in an economically plausible way. Taken together, these results indicate that search-based attention and news-based measures can complement more traditional approaches to monitoring sovereign financial stress in Brazil, while also suggesting that construction choices matter for the final properties of the index.

The remainder of the paper is organized as follows. Section 2 describes the data, the construction of the financial stress indices, and the empirical framework. Section 3 presents the empirical results. Section 4 concludes.

2. Methodology

This section describes the data, the construction of the three financial stress indices, and the empirical framework used to evaluate their relationship with sovereign credit risk. The empirical strategy follows the three goals stated above. We first construct alternative measures of financial stress. We then examine whether those measures contain information about subsequent CDS movements. Finally, we compare how different construction methods perform in practice.

2.1 Data Sources and Sample Period

Our main analysis covers February 2008 to November 2025, yielding 214 monthly observations after alignment across data sources. This period includes several episodes of pronounced stress for Brazilian financial markets, such as the Global Financial Crisis, the 2015–2016 domestic recession and political crisis, the COVID-19 shock, and subsequent periods of fiscal and political uncertainty.

We use three primary data sources. First, Google Trends provides weekly Search Volume Index (SVI) series for Portuguese-language queries related to financial stress. These series are obtained by inserting researcher-defined keywords into Google Trends for Brazil and retrieving the corresponding search intensity measures. The SVI is normalized on a 0–100 scale, where 100 denotes the peak search interest for a term within the selected period and geographic unit. Second, Brazil's 5-year sovereign Credit Default Swap (CDS) spread, obtained from Investing.com, serves as our measure of sovereign credit risk. Higher CDS spreads indicate higher market-implied compensation for default risk. Third, daily IBOVESPA data from Yahoo Finance are used to compute 21-day realized volatility, which provides a market-based benchmark for periods of elevated stress. All series are aggregated to monthly frequency to reduce noise and ensure temporal comparability.

2.2 Construction of the Financial Stress Indices

We construct three methodologically distinct indices; each one is intended to capture financial stress from a different informational angle. The search-based measures are closer to changes in attention and information demand, in line with the interpretation of search intensity advanced by Da et al. (2011a) and Da et al. (2015), while the news-based measure captures changes in the tone and incidence of professional reporting, following the broader logic of text-based uncertainty measures such as Baker et al. (2016) and related work on news and expectations (Candia et al., 2020; Chahrour et al., 2025; Born et al., 2023; Song and Tang, 2018).

The first index is dictionary-based and follows the logic of Da et al. (2011b)as we select 25 Portuguese-language search queries related to financial distress in the Brazilian context. The keywords are not generated mechanically. They are chosen by the researchers from a consolidated literature on financial stress, uncertainty, and investor attention (Da et al., 2011b,a; Castelnuovo and Tran, 2017; Pratap and Priyaranjan, 2023; Baker et al., 2016), and then adapted to Portuguese and to the Brazilian institutional setting. The terms are organized into five tiers. The first tier contains direct crisis language, such as "crise financeira" and "crash bolsa". The second tier covers market stress more broadly, with terms such as "queda ibovespa" and "volatilidade bolsa". The remaining tiers include queries linked to sovereign risk, exchange-rate concerns, banking stress, and investor sentiment. The index is computed as a weighted average of search intensities:

$$\mathrm{FSI}_t^{\mathrm{Dict}} = \frac{\sum_{i=1}^{N} w_i \cdot \mathrm{SVI}_{i,t}}{\sum_{i=1}^{N} w_i}, \tag{1}$$

where $\mathrm{SVI}_{i,t}$ denotes the search intensity for query i in period t, and w_i is the tier-specific weight. We assign weight 1.5 to Tier 1, weight 1.2 to Tier 2, and weight 1.0 to the remaining tiers. The resulting series is normalized to the $[0,1]$ interval using min-max scaling.

The second index is based on machine learning. Following García (2013), we estimate a LASSO specification that selects search queries with predictive content for future equity returns. This specification is not intended to measure sovereign risk directly. Its purpose is methodological: it asks whether a more data-driven selection rule, anchored in a market manifestation of stress, can generate a search-based indicator that is useful for the broader objective of the paper. In that sense, the ML exercise is closer to the literature that uses asset-price dynamics and textual information jointly to recover latent financial conditions (García, 2013; Baker et al., 2016). Specifically, we regress next-period IBOVESPA returns on current-period search intensities:

$$r_{t+1} = \alpha + \sum_{i=1}^{N} \beta_i \cdot \mathrm{SVI}_{i,t} + \varepsilon_{t+1}, \tag{2}$$

where r_{t+1} is the return between t and $t+1$, and the coefficients β_i are estimated under the usual LASSO penalty. The regularization parameter is selected by cross-validation. The machine-learning index is defined as the negative of the fitted return so that higher values correspond to weaker expected market performance. The rationale is that sharp equity-market deterioration is often observed during stress episodes and can therefore serve as an indirect bridge between search behavior and the broader financial-stress concept studied in the paper, even if it does not coincide exactly with sovereign-risk pricing (Hakkio and Keeton, 2009; Holló et al., 2012; Gaglianone and Areosa, 2016).

The third index is based on the text of financial news articles and follows the broad logic of Baker et al. (2016), while also relating to more recent work that uses news content to study uncertainty, beliefs, and macro-financial conditions (Candia et al., 2020; Chahrour et al., 2025; Born et al., 2023; Song and Tang, 2018; Andres-Escayola et al., 2025b). We analyze 33,881 articles from Folha de São Paulo, Valor Econômico, and G1, spanning June 2002 to December 2025. An article is classified as stress-related when it contains at least one term from a dictionary of 99 financial expressions and at least one term from a combined list of 162 stress-related terms and 289 negative terms. For each qualifying article, we compute the score

$$S_a = \sqrt{F_a \times \left(2 \cdot T_a^{\text{stress}} + T_a^{\text{neg}}\right)}, \tag{3}$$

where F_a is the number of financial terms in article a, T_a^{stress} is the number of stress terms, and T_a^{neg} is the number of negative terms. Stress terms receive additional weight because they are more directly tied to financial disruption. The article-level scores are then aggregated to weekly frequency and normalized to the $[0,1]$ interval.

Finally, we report a combined index that aggregates the three components using inverse-variance weights:

$$\mathrm{FSI}_t^{\mathrm{Comb}} = \frac{\sum_j \sigma_j^{-2} \cdot \mathrm{FSI}_t^j}{\sum_j \sigma_j^{-2}}, \tag{4}$$

where $j \in \{\mathrm{Dict}, \mathrm{ML}, \mathrm{News}\}$ and σ_j^2 is the sample variance of each component. This combined series is reported as a descriptive benchmark rather than as the preferred specification.

2.3 Econometric Framework

We begin with Augmented Dickey-Fuller tests to determine the appropriate transformation of the variables. CDS levels are non-stationary in levels ($p = 0.09$), so we work with log differences. The financial stress indices are stationary in levels and are standardized to mean zero and unit variance for comparability across specifications.

To examine whether the stress indicators contain information about sovereign credit risk, we estimate local projections in the spirit of Jordà (2005). This choice is standard in empirical work interested in tracing dynamic responses without imposing a tight parametric structure on the underlying system. This is the paper's main predictive-content exercise. For each horizon $h = 0, 1, \ldots, 12$, we estimate

$$\sum_{j=0}^{h} \Delta \log(\mathrm{CDS})_{t+j} = \alpha_h + \beta_h \cdot \Delta \mathrm{FSI}_t + \sum_{k=1}^{p} \gamma_{h,k} \cdot \Delta \log(\mathrm{CDS})_{t-k} + \varepsilon_{t+h}, \tag{5}$$

where the dependent variable is the cumulative log change in the CDS spread from t to $t+h$. The coefficient β_h measures the cumulative response of CDS spreads to a one-standard-deviation innovation in the FSI. We include four lags of CDS log returns as controls and use Newey-West standard errors with bandwidth $h+1$ to account for serial correlation induced by overlapping horizons (Newey and West, 1987).

We complement the local projections with Granger-causality tests (Granger, 1969). These tests evaluate whether lagged values of the stress indices help predict CDS movements beyond the information already contained in past CDS values. We also examine the relationship between the indices and realized equity-market volatility. Lag lengths from one to four months are considered in the empirical exercises. These tests are used as complementary evidence on temporal precedence and predictive content rather than as a basis for strong causal claims.

To characterize calm and high-stress periods, we estimate a two-state Markov-switching model following Hamilton (1989). This part of the analysis is linked more directly to the measurement dimension of the paper. If an index is capturing meaningful variation in stress, periods of elevated stress should be associated with a distinct high-stress regime. The model assumes that the FSI follows a regime-dependent process:

$$\mathrm{FSI}_t = \mu_{S_t} + \sigma_{S_t}\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \tag{6}$$

where $S_t \in \{0,1\}$ is an unobserved state variable governed by a first-order Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \tag{7}$$

Both the mean and the variance are allowed to differ across regimes. The model is estimated by maximum likelihood, and we report smoothed probabilities of the high-stress regime.

3. Results

3.1 Descriptive Statistics

Table 1 reports the main descriptive statistics. The dictionary-based index has a mean of 0.252 and a standard deviation of 0.176, indicating substantial time variation in search-based stress. The machine-learning index has a higher mean, 0.623, with comparable dispersion. The news-based index is lower on average and less volatile, which is consistent with the lower frequency of strongly stress-related news in normal times. Its aligned sample is also smaller ($N = 166$), reflecting the more restrictive process of identifying and processing relevant newspaper articles. This difference matters for interpretation: the comparison across methodologies is informative, but it is not based on perfectly symmetric samples. CDS spreads average 192 basis points over the sample and vary considerably across episodes.

In our first validation exercise, we compare the indices with realized IBOVESPA volatility. The dictionary-based index is strongly and positively correlated with volatility ($r = 0.70$, $p < 0.001$), which is consistent with its interpretation as a stress measure and with the broader literature linking search intensity to shifts in investor attention and market conditions (Da et al., 2011a, 2015; Castelnuovo and Tran, 2017). The news-based index also shows a positive, albeit more moderate, correlation ($r = 0.33$, $p < 0.001$), a pattern that is in line with work showing that media content tends to intensify around

Table 1
Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Dictionary FSI	0.252	0.176	0.000	1.000
ML FSI	0.623	0.183	0.000	1.000
News FSI	0.135	0.104	0.000	0.544
Combined FSI	0.438	0.117	0.142	0.779
CDS 5Y (bps)	192.4	76.1	97.3	486.9
IBOV Volatility (%)	22.8	12.2	9.2	96.2

periods of macro-financial strain (Baker et al., 2016; Candia et al., 2020; Andres-Escayola et al., 2025b). By contrast, the machine-learning index is negatively correlated with volatility ($r = -0.21$, $p = 0.002$), suggesting that the queries selected for that specification are not well aligned with realized episodes of financial stress in this setting.

3.2 Crisis Episodes

Figure 1 plots the dictionary-based index around major Brazilian and global stress episodes. The index peaks during the COVID-19 shock in March 2020, reaching 0.76, which coincides with the period of exceptional volatility and repeated circuit breakers in the Brazilian stock market. The 2022 electoral period also shows elevated readings, as do fiscal discussions in 2024. These patterns provide descriptive support for the index as a measure of financial stress, although they should not be read as a formal ranking of crisis severity across episodes. In that sense, the figure plays a role similar to the narrative validation often used in the financial-stress and uncertainty literatures (Hakkio and Keeton, 2009; Holló et al., 2012; Baker et al., 2016).

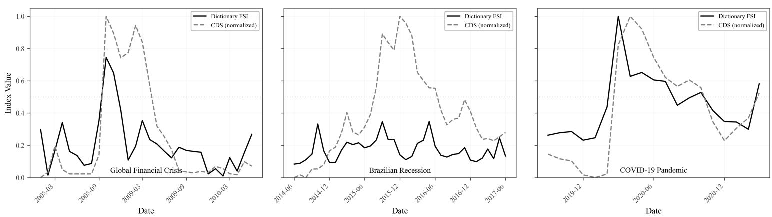

Figure 1
Dictionary-Based FSI During Major Crisis Episodes. Shaded regions indicate known periods of heightened stress. The index reaches its highest value during the COVID-19 shock and rises again during the 2022 electoral period and the 2024 fiscal debate.

The Global Financial Crisis also appears in the series, although with a more moderate peak than later events. One plausible explanation is that internet penetration and the use of online search as a source of financial information were less widespread at the beginning of the sample than in more recent years.

3.3 Predictive Content for Sovereign Credit Risk

Figure 2 presents the cumulative impulse response of CDS spreads to a one-standard-deviation shock in the dictionary-based index. The response is positive throughout the horizon range and remains statistically distinguishable from zero over the full projection window in the baseline specification.

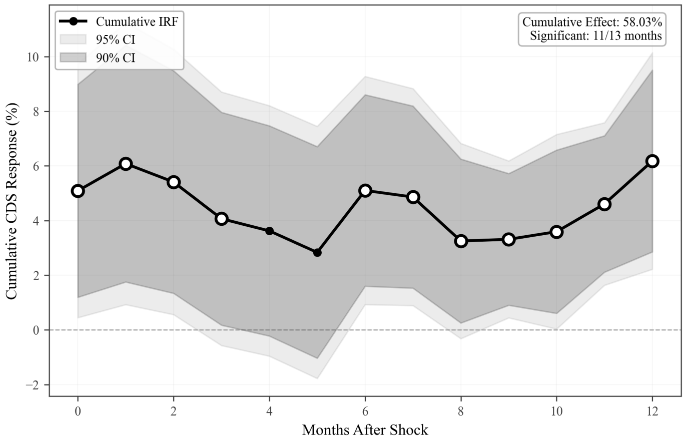

Figure 2

Cumulative Impulse Response of CDS Spreads to a Dictionary-Based FSI Shock. The solid line shows point estimates, and the shaded areas indicate 90% and 95% confidence intervals.

At the 12-month horizon, the cumulative effect reaches 81.45%. We interpret this result as evidence of a sizeable and persistent in-sample association between search-based stress and subsequent sovereign-risk repricing, in line with the view that stress indicators may contain forward-looking information about deteriorating financial conditions (Hakkio and Keeton, 2009; Holló et al., 2012; Gaglianone and Areosa, 2016). In substantive terms, periods of stronger stress-related search intensity tend to be followed by broader CDS widening over the subsequent year. The result is informative about predictive content in the sample, but it is not by itself sufficient to support a structural interpretation of CDS dynamics.

3.4 Comparison Across Construction Methods

Table 2 and Figure 3 compare the four index constructions. In this application, the dictionary-based specification performs better than the alternatives considered here. It is the only measure that delivers a uniformly positive and persistent response pattern over the full horizon range. At the same time, this comparison should be interpreted with some caution because the news-based measure is available over a shorter aligned sample.

Table 2
IRF Results by Methodology

Method	Cumulative IRF (%)	Significant Months
Dictionary FSI	+81.45	13/13
ML FSI	−37.23	2/13
News FSI	+34.39	2/13
Combined FSI	+33.44	3/13

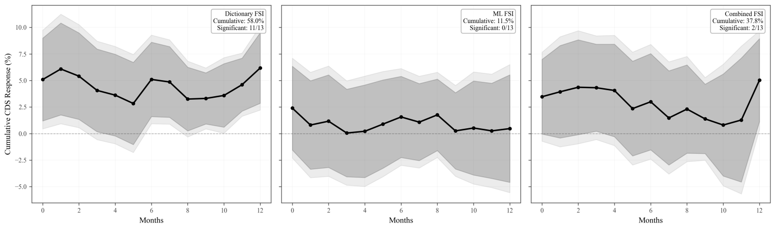

Figure 3
Comparison of Impulse Responses Across FSI Specifications. The dictionary-based measure shows the most stable positive response. The machine-learning index moves in the opposite direction in this application, while the combined index is notably weaker than the dictionary-based series alone.

The machine-learning index produces a negative cumulative response and is statistically significant at only two horizons. In this application, that result is difficult to reconcile with the interpretation of the index as a broad stress measure and is consistent with its negative correlation with realized volatility. More broadly, it suggests that a data-driven procedure targeted at one market margin need not recover the same signal as a literature-based stress dictionary targeted at broader financial conditions (García, 2013; Baker et al., 2016). The news-based index yields a positive cumulative response, but one that is smaller and less precisely estimated. The combined index remains positive, yet its performance is weaker than that of the dictionary-based index. In this setting, the aggregation step appears to dilute rather than sharpen the most informative signal, which is consistent with the idea that heterogeneous textual indicators may capture different moments of the information process rather than a single latent object (Candia et al., 2020; Chahrour et al., 2025).

3.5 Granger-Causality Results

Table 3 reports the Granger-causality results. These exercises provide a complementary view of the timing of the predictive relationships and should be read alongside the local-projection estimates rather than as a stand-alone identification strategy.

Table 3
Granger Causality Test Results

Cause	Effect	Min p-value	Significant
Panel A: Dictionary FSI			
Dictionary FSI	CDS (log-returns)	0.065	*
Dictionary FSI	Volatility	<0.0001	***
Dictionary FSI	News FSI	0.004	***
Volatility	Dictionary FSI	0.012	**
Panel B: News FSI			
News FSI	CDS (log-returns)	0.018	**
News FSI	Volatility	0.002	***
Volatility	News FSI	0.001	***
Panel C: ML and Combined FSI			
ML FSI	CDS (log-returns)	0.087	—
Combined FSI	CDS (log-returns)	0.059	*

$*p < 0.10$; $**p < 0.05$; $***p < 0.01$

The dictionary-based index shows marginal predictive content for CDS movements and substantial predictive content for realized volatility. The news-based index performs somewhat better in the CDS Granger-causality exercise and also predicts volatility. The machine-learning and combined indices are comparatively weaker. An additional result of interest is that the dictionary-based index Granger-causes the news-based index, which is consistent with the interpretation that search behavior may move earlier than formal media coverage, a possibility that fits naturally with the distinction between decentralized attention and editorial filtering emphasized in the search- and news-based literatures (Da et al., 2011a; Candia et al., 2020; Chahrour et al., 2025).

These findings should be read alongside, rather than in isolation from, the local-projection results. Taken together, they suggest that different text-based indicators may contain information at different stages of the transmission process from perceptions of stress to market pricing. More generally, the Granger exercises are best viewed as evidence on temporal ordering and predictive content, not as proof of a causal mechanism.

3.6 Co-movement Between the Dictionary Index and CDS

Figure 4 plots the dictionary-based index together with the 5-year CDS spread. The two series display clear co-movement over the sample. In several episodes, increases in the index precede or coincide with periods of CDS widening, notably during the pandemic shock, the 2022 electoral cycle, and episodes of fiscal concern.

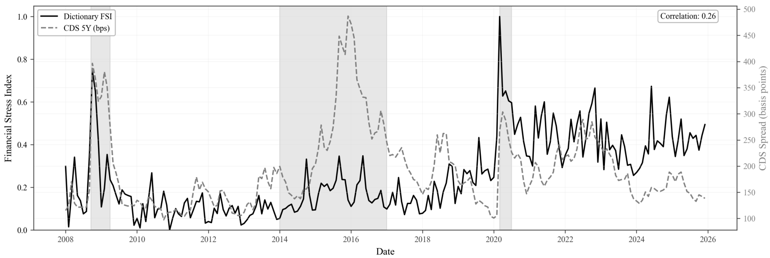

Figure 4

Dictionary-Based FSI and Brazil's 5-Year CDS Spread, 2008–2025. The left axis reports the financial stress index and the right axis the sovereign CDS spread.

This visual pattern is consistent with the formal results and supports the interpretation that the index reflects economically meaningful variation in perceived sovereign risk, in the same broad spirit in which financial-stress indices are often validated against observable market disruptions (Hakkio and Keeton, 2009; Holló et al., 2012; Ferreira and Mattos, 2018).

3.7 Regime Analysis

The Markov-switching estimates identify two distinct regimes. Table 4 summarizes the estimated parameters.

Table 4
Markov-Switching Regime Parameters

Parameter	Calm Regime ($S = 0$)	Crisis Regime ($S = 1$)
Mean (μ)	0.198	0.412
Std. Dev. (σ)	0.089	0.217
Expected Duration (months)	14.2	5.8
Ergodic Probability	0.71	0.29

The high-stress regime is characterized by both a higher mean and substantially higher volatility, which is in line with the stylized fact that stress episodes are associated with more turbulent dynamics and with the use of regime-switching models in the analysis of financial instability (Hamilton, 1989; Monin, 2017). The estimated transition matrix,

$$\hat{\mathbf{P}} = \begin{pmatrix} 0.930 & 0.070 \\ 0.172 & 0.828 \end{pmatrix}, \tag{8}$$

indicates persistent calm periods and shorter, but still persistent, crisis periods.

Figure 5 reports the smoothed probability of the high-stress regime together with the dictionary-based index. The model assigns high stress probabilities to the main episodes of financial disruption in the sample, including the Global Financial Crisis, the Brazilian recession, and the COVID-19 shock.

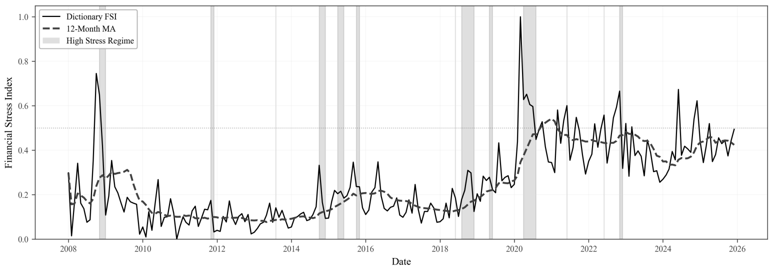

Figure 5

Estimated Financial Stress Regimes, 2008–2025. The shaded areas correspond to months in which the smoothed probability of the high-stress regime exceeds 0.5.

Table 5 summarizes the average probability of the high-stress regime during selected episodes.

Table 5

Illustrative Alignment with Selected Stress Episodes

Episode	Period	Mean $\Pr(S=1)$
Global Financial Crisis	2008–2009	0.78
Brazilian Recession	2014–2016	0.82
COVID-19 Pandemic	2020–2021	0.91
Election Uncertainty	2022	0.74
Fiscal Framework Concerns	2024	0.71

The regime analysis should be interpreted as corroborative rather than definitive evidence. Its role here is close to that of related regime-based exercises in the financial-stress literature, which use latent-state models to distinguish calm from turbulent periods without claiming a fully mechanical identification of crises (Hamilton, 1989; Holló et al., 2012; Monin, 2017). The results indicate that the dictionary-based index tends to move into a distinct high-stress state during periods widely regarded as financially turbulent. This pattern supports the interpretation of that series as a meaningful measure of financial conditions, but it should not be read as a claim of perfect crisis identification.

4. Discussion and Conclusion

The paper has addressed three related questions: how to measure financial stress with text-based data, whether such measures contain information about sovereign CDS dynamics, and how alternative construction methods compare in practice. On the measurement dimension, the dictionary-based search index is the series that most clearly tracks known episodes of turbulence and most closely aligns with realized equity-market volatility. This result is consistent with the use of researcher-selected Google Trends keywords grounded in the consolidated literature on financial stress, uncertainty, and investor attention (Da et al., 2011b; Castelnuovo and Tran, 2017; Pratap and Priyaranjan, 2023; Baker et al., 2016).

On the predictive-content dimension, the local-projection estimates indicate that increases in the dictionary-based index are followed by CDS widening in the sample, while the news-based index also retains useful information, especially in the Granger exercises. This pattern is compatible with a broader literature in which financial-stress and uncertainty measures help summarize information relevant for subsequent market repricing (Hakkio and Keeton, 2009; Holló et al., 2012; Gaglianone and Areosa, 2016; Baker et al., 2016). These findings suggest that text-based indicators can help organize timely information about sovereign stress. At the same time, they should be interpreted in a measured way. The empirical design is informative about association, temporal ordering, and monitoring value, but it is not intended to identify a structural causal mechanism behind sovereign-risk pricing.

The methodological comparison is also informative. In this application, the curated dictionary performs better than the machine-learning alternative. This does not imply that data-driven methods are uninformative in general. Rather, it suggests that an index trained to predict short-run IBOVESPA returns may not coincide with the search measure that best captures broader financial stress relevant for sovereign CDS. That distinction is consistent with the fact that search data, news content, and market prices need not encode the same information at the same time. The weaker performance of the combined index points in the same direction: when component measures differ in timing, noise, and informational content, simple aggregation rules may attenuate rather than improve the final signal.

The news-based index is available on a shorter aligned sample because the identification and processing of newspaper content are more restrictive than the retrieval of search data, so a fully symmetric comparison across all indices is not feasible in the present version. Google Trends data are normalized rather than observed in raw counts, and the empirical validation focuses primarily on one sovereign benchmark. For these reasons, the results should be read as evidence of empirical usefulness rather than as a definitive account of sovereign-risk pricing. Additional robustness analysis remains an important avenue for future work.

With those caveats in mind, the evidence indicates that text-based indicators can complement conventional approaches to measuring financial stress in Brazil, including market-based and composite indicators already used in the literature (Ferreira and Mattos, 2018; Gaglianone and Areosa, 2016; Holló et al., 2012). In this setting, the search-based dictionary index is the most informative of the measures considered here, while the news-based indicator provides an additional, and partly distinct, source of information about sovereign risk. Future research can build on this framework by extending the validation exercises, exploring richer natural-language methods, and considering broader comparative applications across countries in Latin America.

Acknowledgments The authors thank thank financial support from CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico* and CAPES *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*.

Conflict of interest The authors declare no conflict of interest.

Artificial Intelligence The authors used computational tools to assist with coding, data processing, and figure preparation. All outputs were reviewed, revised, and validated by the authors, who remain fully responsible for the content of the manuscript.

Data availability The data used in this study are available from the corresponding author upon reasonable request.

References

Abel, A. B. and Eberly, J. C. (1994). A unified model of investment under uncertainty, *American Economic Review*, *84*(5), 1369–1384.

Andres-Escayola, E., Molina, L., Pérez, J. J. and Vidal, E. (2025a). How economic policy uncertainty spreads across borders: The case of latin america, *Technical Report 2549*, Banco de España.

Andres-Escayola, E., Molina, L., Pérez, J. J. and Vidal, E. (2025b). How economic policy uncertainty spreads across borders: The case of latin america, *Technical Report 2549*, Banco de España.

Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty, *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Batista, A. T. N., Souza, G. H. D., Bressan, V. G. F. and Lamounier, W. M. (2024). The effect of economic policy uncertainty in brazil on corporate investment, *Advances in Scientific and Applied Accounting*, *17*(1), 29–39.

Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment, *The Quarterly Journal of Economics*, *98*(1), 85–106.

Bloom, N. (2007). Uncertainty and the dynamics of R&D, *American Economic Review*, *97*(2), 250–255.

Bloom, N. (2009). The impact of uncertainty shocks, *Econometrica*, *77*(3), 623–685.

Bordo, M. D., Dueker, M. J. and Wheelock, D. C. (2000). Aggregate price shocks and financial instability: An historical analysis, *Technical Report 7652*, National Bureau of Economic Research.

Born, B., Enders, Z. and Müller, G. J. (2023). On fire, news, and expectations, *Technical Report 42*, German Research Foundation's Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour", Humboldt University Berlin.

Candia, B., Coibion, O. and Gorodnichenko, Y. (2020). Communication and the beliefs of economic agents, *Technical Report 27800*, National Bureau of Economic Research.

Castelnuovo, E. and Tran, T. D. (2017). Google it up! a google trends-based uncertainty index for the united states and australia, *Technical Report 6695*, CESifo.

Chahrour, R., Shapiro, A. H. and Wilson, D. (2025). News selection and household inflation expectations, *Technical Report 33837*, National Bureau of Economic Research.

Da, Z., Engelberg, J. and Gao, P. (2011a). In search of attention, *The Journal of Finance*, *66*(5), 1461–1499.

Da, Z., Engelberg, J. and Gao, P. (2011b). In search of attention, *The Journal of Finance*, *66*(5), 1461–1499.

Da, Z., Engelberg, J. and Gao, P. (2015). The sum of all fears: Investor sentiment and asset prices, *The Review of Financial Studies*, *28*(1), 1–32.

Dixit, A. (1992). Investment and hysteresis, *Journal of Economic Perspectives*, *6*(1), 107–132.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms, *The Quarterly Journal of Economics*, *75*(4), 643–669.

Fernandes, M. and Pereira, M. A. P. (2025). Forecasting realized volatility using news flow, *The Quarterly Review of Economics and Finance*, *104*, 102040.

Ferreira, D. M. and Mattos, L. B. d. (2018). Estresses financeiros na economia brasileira durante o regime de metas para inflação, *Economia e Sociedade*, *27*(1), 233–263.

Gaglianone, W. P. and Areosa, W. D. (2016). Financial conditions indicators for brazil, *Technical Report 435*, Central Bank of Brazil.

García, D. (2013). Sentiment during recessions, *The Journal of Finance*, *68*(3), 1267–1300.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, *37*(3), 424–438.

Hakkio, C. S. and Keeton, W. R. (2009). Financial stress: What is it, how can it be measured, and why does it matter?, *Economic Review*, *94*(2), 5–50.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, *57*(2), 357–384.

Holló, D., Kremer, M. and Lo Duca, M. (2012). Ciss – a composite indicator of systemic stress in the financial system, *Technical Report 1426*, European Central Bank.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections, *American Economic Review*, *95*(1), 161–182.

Lefort, B., Benhamou, E., Ohana, J.-J., Saltiel, D., Guez, B. and Jacquot, T. (2024). Stress index strategy enhanced with financial news sentiment analysis for the equity markets.

Louzis, D. P., Vouldis, A. T. and Metaxas, V. L. (2012). Macroeconomic and bank-specific determinants of non-performing loans in greece: A comparative study of mortgage, business and consumer loan portfolios, *Journal of Banking & Finance*, *36*(4), 1012–1027.

Maćkowiak, B., Matějka, F. and Wiederholt, M. (2023). Rational inattention: A review, *Journal of Economic Literature*, *61*(1), 226–273.

Monin, P. (2017). The ofr financial stress index, *Working Paper 17-04*, Office of Financial Research, U.S. Department of the Treasury.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, *55*(3), 703–708.

Park, C.-Y. and Mercado, R. V. J. (2014). Determinants of financial stress in emerging market economies, *Journal of Banking & Finance*, *45*(C), 199–224.

Pereira, M. M. d., Rosa, T. G. d. and Bender Filho, R. (2020). Influência do google trends em ações listadas na bolsa de valores brasileira: evidências a partir da modelagem PVAR, *REAd. Revista Eletrônica de Administração (Porto Alegre)*, *26*(3), 796–818.

Pratap, B. and Priyaranjan, N. (2023). Macroeconomic effects of uncertainty: a google trends-based analysis for india, *Empirical Economics*, *65*(4), 1599–1625.

Püttmann, L. (2018). Patterns of panic: Financial crisis language in historical newspapers, *Ssrn working paper*, SSRN.

Ramos, H. P., Ribeiro, K. K. M. and Perlin, M. S. (2017). The forecasting power of internet search queries in the brazilian financial market, *Mackenzie Management Review (Revista de Administração Mackenzie – RAM)*, *18*(2), 184–210.

Sims, C. A. (2003). Implications of rational inattention, *Journal of Monetary Economics*, *50*(3), 665–690.

Song, D. and Tang, J. (2018). News-driven uncertainty fluctuations, *Technical Report 18-3*, Federal Reserve Bank of Boston.

Vdovychenko, A. and Oros, G. (2014). Financial stress index: estimation and application in empirical researches in ukraine, *MPRA Paper 69337*, Munich Personal RePEc Archive (MPRA).