Name: Vihaan Saxena

Topic: Behavioural Signal Augmentation of Financial Market Stress Detection: A Comparative Multi-System Framework Integrating Alternative Data and Machine Learning

Email: saxenavihaan11@gmail.com

Abstract
Financial markets are conventionally analysed via endogenous indicators such as price dynamics, realised volatility, and trading volume. However, these measures fail to capture the behavioural forces that precipitate and amplify unstable market periods. This study proposes a comparative, multi system analytical framework to evaluate whether behavioural stress indicators, extracted from alternative data sources such as search intensity and information demand can enhance the detection and early identification of financial market stress. By constructing parallel analytical systems that complement traditional market based indicators, the research aims to determine whether behavioural proxies exhibit leading characteristics when it comes to volatility escalation. The findings contribute to the growing field of data driven financial analytics by providing a structured and empirically grounded assessment of behavioural signal integration.

Introduction & Background
Today's evolving financial markets operate at a finely balanced intersection of information processing and collective behavioural response. While the Efficient Market Hypothesis posits that asset prices reflect all possible information regarding volatility, empirical evidence demonstrates consistent deviations from the core of this claim; particularly under conditions of uncertainty (Shiller, 2003). Events such as the 2008 financial crisis and COVID-19 market collapse illustrate how investor sentiment, panic-driven decision-making processes, and herding behaviour can significantly amplify volatility beyond regular theory-backed fundamentals alone (Baker et al., 2020). Traditional indicators, such as the VIX, are inherently reactive and fail to capture the forerunning behavioural responses to market stress (Whaley, 2000). The digitisation of information consumption has enabled the development of scalable proxies for these dynamics, particularly through search engine activity, which display public attention shifts and perceived risk (Preis, Moat, & Stanley, 2013). These signals function in tandem with physiological indicators in neuromarketing. At a population level, it captures aggregate and behavioural stress. Despite their increasing use in institutional finance, such signals are rarely embedded in structured frameworks that systematically evaluate their contribution relative to traditional indicators. This study reframes financial markets not as purely information-efficient systems, but as behaviourally reactive environments where collective attention acts as a leading driver of volatility.

Literature Review
Behavioural finance has established that decision-making under uncertainty deviates from rational models and stems from cognitive biases and emotional responses, thus influencing market outcomes (Kahneman & Tversky, 1979). Volatility indexes such as VIX are used as proxies for market fear; however are completely reflective of preexisting expectations already embedded within the market, inhibiting their ability to act as leading indicators (Whaley, 2000). Recent research has also explored alternative data sources in analysing behavioural dimensions of market activity. Preis et al. (2013) demonstrated that increased search intensity for financially relevant terms precede market declines, suggesting that information seeking behaviour reflects rising uncertainty. Similarly, Da, Engelberg, and Gao (2011) showed that search-based measures of investor attention may explain variations in trading volume and quantitative asset returns. However, these studies are all primarily correlational and do not situate real behavioural signals within controlled comparative frameworks capable of isolating their incremental contribution against market metrics. This study advances literature by introducing a regime-aware, multi-system design that explicitly evaluates behavioural signal integration within a structured analytical architecture. Collectively, these studies demonstrate that behavioural and attention-based signals contain meaningful information about market dynamics; however, they are primarily analysed in isolation rather than within integrated analytical systems.

Research Question

To what extent can behavioural stress indicators derived from alternative data sources optimise the detection, timing and reliability of financial market stress relative to conventional market-based analytical systems?

Methodology

This study implements a reproducible, regime-aware analytical pipeline integrating synchronised financial and behavioural datasets within a controlled comparative structure. Financial data is acquired on a daily basis from publicly available sources (e.g Yahoo Finance, CBOE), including index level prices, log returns and implied volatility measures (VIX) for major benchmarks such as the S&P 500 and NIFTY 50 spanning over a multi year horizon. Series are transformed into standardised returns and rolling volatility estimates use fixed window operators to ensure temporal consistency. Behavioural data is extracted via the Google Trends API (pytrends) for a defined lexicon of stress-related queries ("market crash", "recession," "stock market fear", "stock sell recommendations"), with term selection continuously refined based on signal persistence and event relevance. Search indices are aggregated into a composite behavioural stress factor using empirically calibrated weights derived from historical instances of volatility spikes.

All variables are time-aligned to a common index, standardised, and denoised using rolling smoothing filters. The dataset is then segmented into stable, transitional and high stress regimes using threshold based classification of volatility measure, cross-validated against known macroeconomic events (e.g., COVID-19 market shock). Event windows are constricted to major volatility inflection points to enable consistent cross-system evaluation. Three analytical systems are specified: (i) a baseline system using only market-derived variables; (ii) an augmented system incorporating the behavioural stress factor via feature integration and lag-structured inputs; and (iii) a supervised machine learning system employing ensemble classifiers (e.g. Random Forest, Gradient Boosting) trained on combined feature sets. This learning pipeline uses rolling-window train-test splits to preserve temporal causality, with hyperparameters optimised via grid search techniques and performance evaluated on out-of-sample segments.

To assess directional informativeness, a lead-lag analysis is conducted by systematically shifting behavioural features relative to market variables and then quantifying alignment with observed volatility escalations. System performance is evaluated across detection accuracy, early warning lead time, signal stability across regimes and false positive incidence. Classification outputs will be assessed via precision, recall and F1-score. Results are visualised through time-series overlays, event-aligned response plots and feature importance rankings to ensure cohesiveness and interpretability and direct cross-system comparison.

Research Significance

This research is innovative in its formalisation of behavioural signal integration within a controlled, regime aware analytical framework for financial market stress detection, addressing a critical gap in financial models that has been overlooked for far too long. While alternative data sources have been looked into, its structured incorporation into comparative, system level frameworks remains underdeveloped. By integrating financial indicators with behavioural proxies, the scope for overlooked data and inaccurate predictions relating to market analysis decreases drastically. A clear example of this can be seen during the 2021 GameStop short squeeze, where surges in online attention and search activity preceded extreme price volatility, reflecting how herd behaviour can drive markets before traditional indicators have the opportunity to react. By systematically capturing and structuring such signals, this research shifts behavioural data from background noise to a core tool for anticipating financial stress and enhancing the understanding of the world of market econometrics.

Bibliography (APA 7th Edition)

Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). COVID-induced economic uncertainty. *National Bureau of Economic Research*. https://doi.org/10.3386/w26983

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343. https://doi.org/10.1016/S0304-405X(98)00027-0

Chen, M., Yin, H., Huang, Y., & Zhu, Z. (2025). Research into the factors that cause a price bubble in GameStop during 2021. *Advances in Economics, Management and Political Sciences*, *202*(1), 262–270. https://doi.org/10.54254/2754-1169/2024.25133

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Heston, S. L., & Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, *73*(3), 67–83. https://doi.org/10.2469/faj.v73.n3.3

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684. https://doi.org/10.1038/srep01684

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, *17*(1), 83–104. https://doi.org/10.1257/089533003321164967

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Whaley, R. E. (2000). The investor fear gauge. *Journal of Portfolio Management*, *26*(3), 12–17. https://doi.org/10.3905/jpm.2000.319728

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter "I hope it is not as bad as I fear." *Procedia - Social and Behavioral Sciences*, *26*, 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562