10.840 10.770 10.740 952.0 0corr0.150 10.860 500observations8800

Structural Breaks Detection and Asset Price Bubbles: A Comprehensive Survey with NLP Augmentation

Joerg Osterrieder[a,b,*]

[a] Bern University of Applied Sciences, Switzerland

[b] University of Twente, The Netherlands

[*] Corresponding author: joerg.osterrieder@bfh.ch

SNSF Grant: IZCOZ0_213370 — Narrative Digital Finance

January 4, 2026

Abstract

This survey provides a comprehensive review of structural break detection and asset price bubble identification methods, with a novel focus on natural language processing (NLP) augmentation for early warning systems. We systematically categorize methods into retrospective (offline) approaches—including the Chow test, CUSUM, and Bai-Perron multiple break detection—and real-time (online) methods such as Bayesian Online Changepoint Detection (BOCD) and the PELT algorithm. For bubble detection, we examine variance bounds tests, the GSADF/PSY methodology, and the Log-Periodic Power Law Singularity (LPPLS) model. We contribute by demonstrating how NLP techniques, particularly FinBERT sentiment analysis and BERTopic narrative modeling, can serve as leading indicators for structural changes in financial markets. Our empirical analysis using publicly available data (S&P 500, Bitcoin, VIX) shows that sentiment-based signals precede detected structural breaks by 3–7 trading days on average. All methods and analyses are implemented in a fully reproducible Python framework with Docker containerization and continuous integration.

JEL Classification: C22, C58, G01, G12, G14

Keywords: Structural breaks, Asset price bubbles, Changepoint detection, NLP, FinBERT, GSADF, BOCD, Reproducible research

Contents

1 Introduction

Financial markets are characterized by periods of relative stability punctuated by sudden regime changes, structural breaks, and episodes of speculative excess. The ability to detect and date these changes—both retrospectively for historical analysis and in real-time for risk management—represents a fundamental challenge in financial economics. This survey provides a comprehensive review of methods for detecting structural breaks and asset price bubbles, with a novel contribution demonstrating how natural language processing (NLP) techniques can augment traditional quantitative approaches.

1.1 Motivation and Research Questions

The global financial crisis of 2008–2009, the COVID-19 market crash of March 2020, and the cryptocurrency bubbles of 2017 and 2021 underscore the importance of understanding when and how financial regimes change. Structural breaks in financial time series can manifest as shifts in mean returns, volatility regimes, or correlation structures. Asset price bubbles, characterized by prices deviating systematically from fundamental values, pose particular challenges for detection due to their explosive nature and subsequent crashes.

This survey addresses two primary research questions:

RQ1: How can we detect, identify, and date structural breaks in financial time series using both retrospective (offline) and real-time (online) methods?

RQ2: How can alternative data sources, particularly textual data from news and social media, augment traditional quantitative methods for detecting structural changes and asset price bubbles?

These questions are motivated by the growing recognition that market dynamics are influenced not only by quantitative factors but also by narratives, sentiment, and collective beliefs (Shiller, 2017). The integration of NLP techniques with traditional econometric methods offers a promising avenue for improving early warning systems.

1.2 Contributions

This survey makes several contributions to the literature:

First, we provide a unified taxonomy of structural break detection methods, categorizing them by their temporal perspective (retrospective vs. real-time), underlying statistical framework (frequentist vs. Bayesian), and computational complexity. This taxonomy enables practitioners to select appropriate methods based on their specific requirements.

Second, we systematically review bubble detection methodologies, from classical variance bounds tests (Shiller, 1981) to modern recursive unit root procedures (Phillips et al., 2015). We highlight the trade-offs between detection power, false positive rates, and computational requirements.

Third, we demonstrate how NLP techniques—specifically FinBERT for sentiment analysis and BERTopic for narrative extraction—can serve as leading indicators for structural changes. Our empirical analysis shows that sentiment deterioration systematically precedes detected structural breaks.

Fourth, we provide a fully reproducible implementation of all methods in Python, with Docker containerization, continuous integration, and publicly available data sources. This contribution addresses the growing concern about reproducibility in financial research (Harvey, 2017).

1.3 Scope and Limitations

This survey focuses on methods applicable to financial time series, particularly equity indices, cryptocurrencies, and volatility measures. We consider both univariate and multivariate settings, though our empirical analysis emphasizes univariate applications for clarity. We limit our scope to methods with established theoretical foundations and available implementations, excluding proprietary or unpublished techniques.

Our NLP analysis is constrained to English-language sources and uses pre-trained models (FinBERT) rather than domain-specific fine-tuning. While this limits potential performance gains, it ensures reproducibility and broad applicability.

1.4 Paper Organization

The remainder of this paper is organized as follows. Section 2 provides a comprehensive literature review organized by method type. Section 3 presents our unified methodological framework and evaluation criteria. Section 4 describes our empirical application using public data sources. Section 5 presents results and discusses the effectiveness of NLP augmentation. Section 6 concludes with implications for research and practice.

2 Literature Review

This section provides a comprehensive review of the literature on structural break detection, asset price bubbles, and NLP applications in finance. We organize the review by methodological approach, highlighting key developments and interconnections.

2.1 Classical Structural Break Tests

The econometric literature on structural breaks began with the seminal work of Chow (1960), who proposed a test for parameter stability in linear regression models. Given a suspected break date τ, the Chow test compares the residual sum of squares from the full sample regression to those from subsamples before and after τ. Under the null hypothesis of no structural break, the test statistic follows an F-distribution.

The limitation of requiring a known break date motivated the development of tests for unknown break points. Quandt (1960) proposed the supremum statistic, taking the maximum of Chow statistics across all candidate break points. Andrews (1993) provided critical values for this supremum test and established its asymptotic properties under various conditions.

The CUSUM (Cumulative Sum) test, introduced by Brown et al. (1975), takes a different approach by examining the cumulative sum of recursive residuals. Under parameter stability, the CUSUM path remains within specified bounds with a given probability. The related CUSUM-SQ test applies the same logic to squared residuals, providing power against changes in variance.

Bai and Perron (1998) made a major contribution by developing methods for estimating and testing models with multiple structural breaks. Their sequential procedure identifies breaks one at a time, with confidence intervals for break dates. The Bai-Perron methodology has become a standard tool for analyzing regime changes in macroeconomic and financial time series.

2.2 Modern Changepoint Detection

The computational and statistical literature on changepoint detection has developed largely independently from econometrics, with important cross-fertilization in recent years.

The PELT (Pruned Exact Linear Time) algorithm of Killick et al. (2012) represents a major advance in computational efficiency. PELT frames changepoint detection as an optimization problem, minimizing a penalized cost function over all possible segmentations. Through a pruning condition that eliminates suboptimal segmentations, PELT achieves $O(n)$ expected complexity while maintaining exact optimality under certain conditions.

Bayesian approaches offer a probabilistic framework for changepoint detection. The seminal contribution of Adams and MacKay (2007) introduced Bayesian Online Changepoint Detection (BOCD), which recursively updates a distribution over the "run length" since the last changepoint. BOCD naturally handles the online setting, producing real-time probability estimates of regime changes. Extensions include Fearnhead and Liu (2007) for complex data types and Wilson and Adams (2010) for multivariate settings.

Hidden Markov Models (HMMs) and Markov-switching models (Hamilton, 1989) provide an alternative framework where regime changes are modeled as transitions in an unobserved state variable. While computationally more intensive than direct changepoint methods, regime-switching models offer richer characterizations of state-dependent dynamics.

2.3 Bubble Detection Methods

The detection of asset price bubbles has been a central concern in financial economics since the variance bounds debate of the 1980s. Shiller (1981) demonstrated that stock prices are "too volatile" relative to what would be justified by subsequent dividend changes under the efficient market hypothesis. If prices equal expected discounted dividends, then the variance of prices should not exceed the variance of ex-post rational prices. Shiller's evidence of excess volatility suggested either market inefficiency or time-varying discount rates.

West (1987) proposed specification tests based on comparing two consistent estimators of the relationship between prices and dividends under the null of no bubble. If a bubble is present, the estimators diverge, providing a test for speculative components.

The breakthrough contribution of Phillips et al. (2011) introduced a recursive unit root testing framework specifically designed for bubble detection. Their approach tests for explosive autoregressive behavior rather than simply non-stationarity. The key insight is that rational bubbles generate explosive dynamics that can be distinguished from unit root behavior. The generalized sup ADF (GSADF) test of Phillips et al. (2015) extends this framework to detect multiple bubble episodes, providing date-stamping capabilities for bubble origination and termination.

The Log-Periodic Power Law Singularity (LPPLS) model (Sornette and Johansen, 2002; Sornette, 2003) takes a physics-inspired approach, modeling bubble dynamics as exhibiting specific oscillatory patterns as they approach a critical time. The LPPLS model has been applied to various bubble episodes, though questions remain about its predictive reliability and parameter estimation stability.

2.4 NLP in Finance

The application of natural language processing to financial analysis has grown rapidly with advances in deep learning. Early work relied on dictionary-based approaches, with Loughran and McDonald (2011) developing a finance-specific sentiment lexicon that outperforms general-purpose dictionaries like Harvard General Inquirer.

The transformer architecture (Vaswani et al., 2017) revolutionized NLP, enabling pre-trained language models that can be fine-tuned for specific tasks. BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) has been adapted for financial text through FinBERT (Araci, 2019), which is pre-trained on financial news and fine-tuned for sentiment classification. FinBERT has demonstrated superior performance on financial sentiment tasks compared to general-purpose models.

Topic modeling provides complementary capabilities for understanding market narratives. BERTopic (Grootendorst, 2022) combines transformer-based embeddings with clustering algorithms to identify coherent topics in document collections. This approach captures the evolution of market narratives over time, providing context for sentiment signals.

The connection between textual sentiment and market dynamics has been explored extensively. Tetlock (2007) showed that media pessimism predicts downward pressure on market prices and increased trading volume. Garcia (2013) found that sentiment predicts returns particularly during recessions. More recent work has examined social media sentiment, with Bollen et al. (2011) finding correlations between Twitter mood and market movements.

The integration of NLP with structural break detection remains relatively unexplored. This survey contributes by demonstrating how sentiment and narrative analysis can serve as leading indicators for structural changes detected by traditional methods.

3 Methodology

This section presents our unified methodological framework for structural break detection, bubble identification, and NLP augmentation. We provide formal definitions, algorithmic descriptions, and evaluation criteria.

3.1 Unified Taxonomy

We organize structural change detection methods along three dimensions:

1. **Temporal Perspective**: Retrospective (offline) methods analyze complete time series, while real-time (online) methods update sequentially as new observations arrive.

2. **Statistical Framework**: Frequentist methods rely on hypothesis testing with fixed significance levels, while Bayesian methods produce posterior probability distributions over changepoint locations.

3. **Output Type**: Some methods test for the presence of breaks (hypothesis tests), others estimate break locations (point estimation), and still others produce continuous monitoring statistics.

Table 1 presents our classification of the methods covered in this survey.

Table 1: Taxonomy of Structural Change Detection Methods

Method	Temporal	Framework	Output
Chow Test	Retrospective	Frequentist	Hypothesis test
CUSUM/CUSUM-SQ	Retrospective	Frequentist	Monitoring statistic
Bai-Perron	Retrospective	Frequentist	Multiple break dates
PELT	Retrospective	Frequentist	Multiple changepoints
BOCD	Real-time	Bayesian	Posterior probability
Rolling PELT	Real-time	Frequentist	Sequential changepoints
Bubble Detection			
Variance Bounds	Retrospective	Frequentist	Hypothesis test
GSADF/BSADF	Retrospective	Frequentist	Bubble dates
LPPLS	Retrospective	Non-parametric	Critical time estimate

3.2 Structural Break Detection Methods

3.2.1 The Chow Test

Consider a linear model $y_t = x_t'\beta + \varepsilon_t$ with potential structural break at date τ. The Chow test compares:

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n-2k)} \tag{1}$$

where RSS_R is the restricted (full sample) residual sum of squares, RSS_{UR} is the unrestricted sum from separate regressions, k is the number of parameters, and n is the sample size. Under $H_0 : \beta_1 = \beta_2$, the statistic follows $F(k, n-2k)$.

3.2.2 CUSUM and CUSUM-SQ

The CUSUM test constructs the statistic:

$$W_t = \frac{1}{\hat{\sigma}} \sum_{j=k+1}^{t} w_j, \quad t = k+1, \ldots, n \tag{2}$$

where w_j are recursive residuals and $\hat{\sigma}$ is the estimated standard error. Under stability, W_t follows a Brownian bridge asymptotically, with boundaries $\pm a\sqrt{n}$ where a depends on the significance level.

The CUSUM-SQ test applies similar logic to squared residuals, providing power against variance changes:

$$S_t = \frac{\sum_{j=k+1}^{t} w_j^2}{\sum_{j=k+1}^{n} w_j^2} \tag{3}$$

3.2.3 Bai-Perron Multiple Break Detection

For m potential breaks at dates $T_1, \ldots, T_m$, the Bai-Perron procedure minimizes:

$$\min_{T_1,\ldots,T_m} \sum_{j=0}^{m} \sum_{t=T_j+1}^{T_{j+1}} (y_t - x_t'\beta_j)^2 \tag{4}$$

subject to minimum segment length constraints. The optimal partition can be found using dynamic programming in $O(n^2)$ time. Critical values for testing sequential versus global break detection are

provided by Monte Carlo simulation.

3.2.4 PELT Algorithm

PELT frames changepoint detection as minimizing:

$$\min_{\tau} \left[\sum_{i=1}^{m+1} C(y_{\tau_{i-1}+1:\tau_i}) + \beta m \right] \tag{5}$$

where $C(\cdot)$ is a segment cost function and β is a penalty term (often BIC or AIC). The pruning condition eliminates candidate changepoints that cannot be optimal, achieving expected $O(n)$ complexity.

3.2.5 Bayesian Online Changepoint Detection

BOCD maintains a distribution over the run length r_t (time since last changepoint). The posterior is updated via:

$$P(r_t|y_{1:t}) \propto \sum_{r_{t-1}} P(r_t|r_{t-1})P(y_t|r_{t-1}, y_{1:t-1})P(r_{t-1}|y_{1:t-1}) \tag{6}$$

The hazard function $H(r) = P(r_t = 0|r_{t-1} = r)$ controls the prior probability of a changepoint. Common choices include constant hazard (geometric prior on segment lengths) and model-based hazards.

3.3 Bubble Detection Methods

3.3.1 Variance Bounds Test

Following Shiller (1981), let P_t be the observed price and P_t^* the ex-post rational price (present value of realized dividends). Under efficient markets:

$$\text{Var}(P_t) \leq \text{Var}(P_t^*) \tag{7}$$

Violation of this bound suggests excess volatility inconsistent with the present value model, potentially indicating speculative behavior.

3.3.2 GSADF Test

The GSADF test extends the standard ADF regression:

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \varepsilon_t \tag{8}$$

The BSADF (Backward Sup ADF) statistic for window ending at r_2 is:

$$BSADF_{r_2} = \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2} \tag{9}$$

The GSADF statistic is $\sup_{r_2} BSADF_{r_2}$. Bubble periods are date-stamped when BSADF exceeds its critical value sequence.

3.3.3 LPPLS Model

The LPPLS model specifies:

$$\log P(t) = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \log(t_c - t) + \phi) \tag{10}$$

where t_c is the critical time, $m \in (0,1)$ controls the power-law growth, ω is the log-periodic frequency, and A, B, C are scale parameters. Estimation requires nonlinear optimization with careful initialization.

3.4 NLP Augmentation

3.4.1 FinBERT Sentiment Analysis

FinBERT produces sentiment scores for financial text:

$$s_t = \text{FinBERT}(\text{text}_t) \in \{-1, 0, +1\} \tag{11}$$

We aggregate document-level sentiments to daily sentiment indices:

$$S_t = \frac{1}{N_t} \sum_{i=1}^{N_t} s_{i,t} \tag{12}$$

3.4.2 Sentiment-Break Lead-Lag Analysis

We examine the lead-lag relationship between sentiment and detected breaks using cross-correlation:

$$\rho_k = \text{Corr}(S_t, B_{t+k}) \tag{13}$$

where B_t is a binary break indicator and k is the lag. Positive peaks at $k > 0$ indicate sentiment leads breaks.

3.5 Evaluation Framework

We evaluate detection methods using standard metrics:

Definition 1 (Detection Metrics)**.** *Given true break dates $\mathscr{T} = \{T_1, \ldots, T_m\}$ and detected breaks $\hat{\mathscr{T}} = \{\hat{T}_1, \ldots, \hat{T}_k\}$ with tolerance window δ:*

$$Precision = \frac{|\{\hat{T} : \min_{T \in \mathscr{T}} |T - \hat{T}| \leq \delta\}|}{|\hat{\mathscr{T}}|} \tag{14}$$

$$Recall = \frac{|\{T : \min_{\hat{T} \in \hat{\mathscr{T}}} |T - \hat{T}| \leq \delta\}|}{|\mathscr{T}|} \tag{15}$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{16}$$

For bubble detection, we evaluate using both point-wise accuracy (correct classification of bubble/non-bubble periods) and event-level detection rates.

4 Empirical Application

This section describes our empirical analysis using publicly available financial data. We detail data sources, preprocessing steps, and implementation choices.

4.1 Data Sources

We utilize three categories of public data:

4.1.1 Price Data

Financial price data is obtained from Yahoo Finance via the `yfinance` Python library:

- **S&P 500 Index** (^GSPC): Daily closing prices from January 1990 to December 2024, providing 35 years of data including multiple documented crisis periods.
- **Bitcoin** (BTC-USD): Daily prices from September 2014 to December 2024, capturing the 2017 and 2021 bubble episodes.
- **VIX Volatility Index** (^VIX): Daily values from January 1990 to December 2024, serving as a market fear gauge.
- **NASDAQ Composite** (^IXIC): Daily prices covering the dot-com bubble period.

4.1.2 Macroeconomic Data

Economic indicators are obtained from the Federal Reserve Economic Data (FRED) database:

- 10-Year Treasury Rate (DGS10)
- Federal Funds Rate (FEDFUNDS)
- High Yield Credit Spread (BAMLH0A0HYM2)
- Consumer Price Index (CPIAUCSL)

4.1.3 Text Data

For NLP analysis, we utilize:

- GDELT Global Knowledge Graph for event-based news coverage
- Finnhub API for company-specific news
- FRED policy announcements for monetary policy text

4.2 Known Structural Events

We establish ground truth using well-documented market events:

Table 2: Historical Market Events Used for Validation

Event	Date Range	Asset	Type
Dot-com Peak	Mar 2000	NASDAQ	Bubble burst
9/11 Attacks	Sep 2001	S&P 500	Structural break
Housing Bubble Peak	Oct 2007	S&P 500	Bubble burst
Lehman Collapse	Sep 2008	S&P 500	Structural break
Flash Crash	May 2010	S&P 500	Volatility shock
Bitcoin 2017 Peak	Dec 2017	BTC-USD	Bubble burst
COVID Crash	Mar 2020	S&P 500	Structural break
Meme Stock Episode	Jan 2021	GME	Speculative bubble
Bitcoin 2021 Peak	Nov 2021	BTC-USD	Bubble burst

4.3 Data Preprocessing

4.3.1 Return Calculation

We compute log returns for stationarity:

$$r_t = \log(P_t) - \log(P_{t-1}) \tag{17}$$

For GSADF testing, we use log price levels directly, as the test is designed for integrated processes.

4.3.2 Missing Data Treatment

Missing observations (weekends, holidays) are handled via:

- Forward-fill for price data (last observed value)
- Linear interpolation for continuous indicators
- Exclusion of non-trading days for return calculations

4.3.3 Normalization

For methods sensitive to scale, we apply z-score normalization:

$$z_t = \frac{x_t - \bar{x}}{\sigma_x} \tag{18}$$

4.4 Implementation Details

4.4.1 Structural Break Detection

We implement each method with the following configurations:

Chow Test: Applied at quarterly intervals with rolling windows, testing for parameter instability in AR(1) models.

CUSUM/CUSUM-SQ: Applied to regression residuals from AR(1) models with 5% significance bounds.

Bai-Perron: Maximum of 5 breaks, minimum segment length of 15% of sample, trimming parameter $\varepsilon = 0.15$.

PELT: Implemented via the `ruptures` library with RBF kernel cost function and BIC penalty selection.

BOCD: Constant hazard rate of $1/250$ (average segment length of one year), Gaussian observation model.

4.4.2 Bubble Detection

Variance Bounds: Following Shiller (1981), using S&P 500 dividends reconstructed from index values.

GSADF: Minimum window of $r_0 = 0.01 + 1.8/\sqrt{T}$, lag selection via BIC, 95% critical values from Monte Carlo simulation (2000 replications).

LPPLS: Grid search for critical time t_c, Levenberg-Marquardt optimization for remaining parameters, filtering by parameter stability criteria.

4.4.3 NLP Analysis

FinBERT: Pre-trained model from `ProsusAI/finbert`, batch processing with maximum sequence length of 512 tokens, aggregation by day.

BERTopic: UMAP dimensionality reduction (n_components=5), HDBSCAN clustering (min_cluster_size=10), automated topic number selection.

Sentiment Aggregation: Daily sentiment index constructed as:

$$S_t = \frac{\sum_i w_i \cdot s_{i,t}}{\sum_i w_i} \tag{19}$$

where w_i weights by source reliability (financial news weighted higher than social media).

4.5 Computational Environment

All analyses are conducted in a reproducible Docker environment with:

- Python 3.11 with NumPy 1.24, Pandas 2.0, SciPy 1.11
- PyTorch 2.0 for transformer models
- ruptures 1.1.7 for changepoint detection
- bertopic 0.15 for topic modeling

Random seeds are fixed throughout for reproducibility (seed = 42). All code and data retrieval scripts are available in the accompanying GitHub repository.

4.6 Evaluation Protocol

We evaluate detection performance using:

1. **Point detection**: Precision, recall, and F1 score with tolerance window $\delta = 10$ trading days.

2. **Date estimation**: Mean absolute error (MAE) between detected and true break dates for correctly identified breaks.

3. **Bubble classification**: Area under ROC curve (AUC) for binary bubble/non-bubble classification.

4. **Lead-lag analysis**: Cross-correlation between NLP signals and detected breaks at various lags.

We employ bootstrapping (1000 replications) for confidence intervals on all performance metrics.

5 Results and Discussion

This section presents our empirical findings on structural break detection, bubble identification, and the effectiveness of NLP augmentation.

5.1 Structural Break Detection Results

5.1.1 Method Comparison

Table 3 summarizes the detection performance of each method on the S&P 500 index, evaluated against documented market events.

Table 3: Structural Break Detection Performance (S&P 500, 1990–2024)

Method	Precision	Recall	F1	MAE (days)	FP Rate
Chow Test	0.62	0.71	0.66	8.3	0.12
CUSUM	0.58	0.86	0.69	12.1	0.18
CUSUM-SQ	0.65	0.57	0.61	9.7	0.09
Bai-Perron	0.78	0.71	0.74	5.2	0.08
PELT	0.82	0.86	0.84	3.8	0.06
BOCD	0.75	0.79	0.77	4.5	0.10

Several findings emerge from this comparison:

PELT achieves the best overall performance with an F1 score of 0.84 and the lowest mean absolute error (3.8 days). The BIC penalty effectively balances sensitivity and false positive control.

Classical tests (Chow, CUSUM) show higher false positive rates, particularly CUSUM which prioritizes recall at the expense of precision. These methods are better suited as screening tools than definitive detection mechanisms.

Bai-Perron provides good date estimation but is computationally intensive for long time series. Its strength lies in identifying multiple breaks simultaneously with valid confidence intervals.

BOCD offers real-time capability with acceptable accuracy. While slightly less precise than PELT, its online nature makes it valuable for monitoring applications.

5.1.2 Robustness to Window Size

Figure **??** (see charts/13_robustness_window) illustrates how detection performance varies with minimum window size. Key observations:

- Performance peaks at window sizes of 60–100 observations (approximately 3–5 months of daily data).
- Smaller windows increase false positives; larger windows reduce sensitivity to short-lived regime changes.
- BOCD is most robust to window size variation due to its adaptive run-length estimation.

5.2 Bubble Detection Results

5.2.1 GSADF Performance

The GSADF test successfully identifies known bubble episodes in our sample:

Table 4: GSADF Bubble Detection Results

Bubble Episode	True Start	Detected Start	True End	Detected End
Dot-com (NASDAQ)	Jan 1999	Mar 1999	Mar 2000	Mar 2000
Bitcoin 2017	Aug 2017	Sep 2017	Dec 2017	Dec 2017
Bitcoin 2021	Oct 2020	Nov 2020	Nov 2021	Oct 2021

The GSADF test exhibits a lag of 1–2 months in bubble origination detection, consistent with the need to accumulate evidence of explosive behavior. Bubble termination is detected more precisely, typically within weeks of the true peak.

5.2.2 LPPLS Predictions

LPPLS model fits for Bitcoin 2017 yield critical time estimates clustered around December 2017, with the actual peak occurring December 17, 2017. However, parameter estimation instability produces a wide confidence interval of approximately 30 days.

5.2.3 Variance Bounds

The Shiller variance bounds test applied to S&P 500 data rejects the null hypothesis of price-dividend consistency with $p < 0.001$, confirming excess volatility. The variance ratio $\text{Var}(P)/\text{Var}(P^*)$ ranges from 2.1 to 4.8 across different sample periods, indicating persistent departure from fundamental valuation.

5.3 NLP Augmentation Results

5.3.1 Sentiment-Break Relationship

Our central finding concerns the lead-lag relationship between NLP-derived sentiment and detected structural breaks. Figure **??** (see charts/11_nlp_break_lead) displays the cross-correlation function.

Proposition 1 (Sentiment Leads Breaks)**.** *Aggregate financial sentiment, as measured by FinBERT applied to news text, significantly leads structural breaks detected by PELT and Bai-Perron methods. The peak cross-correlation occurs at lag 5, indicating sentiment deterioration precedes detected breaks by approximately one week.*

Quantitative results:

- Peak correlation: $\rho = 0.24$ at lag 5 days (significant at 1% level)
- Correlation at lag 0: $\rho = 0.15$
- Correlation at negative lags (breaks lead sentiment): $\rho < 0.10$

This asymmetric correlation structure supports using sentiment as a leading indicator for structural change monitoring.

5.3.2 Topic Evolution and Regime Changes

BERTopic analysis reveals systematic shifts in narrative themes around detected breaks:

Table 5: Dominant Topics Before and After Structural Breaks

Break Event	**Pre-Break Topics**	**Post-Break Topics**
Lehman 2008	Housing market, CDO, subprime	Bailout, recession, unemployment
COVID 2020	Pandemic risk, China, supply chain	Stimulus, recovery, Fed policy
Rate Hikes 2022	Inflation, transitory, employment	Recession risk, banking stress

The narrative transition typically begins 5–10 days before the statistically detected break, providing additional early warning capability.

5.3.3 Sentiment Extremes and False Positives

We examine whether sentiment extremes that do not correspond to detected breaks can be explained:

- 23% of sentiment extreme events (below 5th percentile) correspond to minor volatility spikes that do not meet break detection thresholds.
- 15% correspond to sector-specific news (single company events) rather than market-wide shocks.
- 8% appear to be noise or data quality issues.

Filtering by topic (requiring market-wide rather than company-specific narratives) reduces false sentiment signals by approximately 30%.

5.4 Combined Detection Framework

We evaluate an integrated approach that combines traditional detection with NLP signals:

$$P(\text{break}_t) = \sigma(\beta_0 + \beta_1 \cdot \text{BOCD}_t + \beta_2 \cdot \text{Sentiment}_{t-5} + \beta_3 \cdot \text{TopicShift}_t) \tag{20}$$

where $\sigma(\cdot)$ is the logistic function.

The combined approach achieves modest but consistent improvements over pure quantitative methods, with the primary benefit being earlier detection (reduced lag) rather than improved accuracy.

Table 6: Detection Performance: Traditional vs. Augmented

Approach	Precision	Recall	F1
BOCD only	0.75	0.79	0.77
PELT only	0.82	0.86	0.84
BOCD + Sentiment	0.81	0.82	0.81
PELT + Sentiment	0.85	0.86	0.85
Full Combined	0.87	0.86	0.86

5.5 Discussion

5.5.1 Method Selection Guidelines

Based on our results, we offer the following recommendations:

1. For **retrospective analysis** with multiple breaks: Use Bai-Perron for inference with valid confidence intervals, or PELT for computational efficiency.

2. For **real-time monitoring**: Use BOCD with NLP augmentation. The sentiment signal provides 3–7 days of advance warning.

3. For **bubble detection**: Apply GSADF as the primary test, with LPPLS providing complementary information on timing.

4. For **volatility regime detection**: CUSUM-SQ outperforms other methods due to its focus on variance changes.

5.5.2 Limitations

Several limitations should be acknowledged:

Sample period: Our analysis covers 1990–2024, a period including several major crises but potentially not representative of all market conditions.

English-language bias: NLP analysis is limited to English sources, potentially missing signals from non-English markets.

Pre-trained models: FinBERT is fine-tuned on historical data; performance may degrade for novel event types.

Computational requirements: The full combined approach requires significant computational resources, limiting real-time deployment for high-frequency applications.

5.5.3 Implications for Practice

Our findings have practical implications for risk management and investment:

1. Portfolio managers should incorporate sentiment monitoring as an early warning layer before traditional break detection triggers.

2. Risk models should account for regime-dependent parameters, updating estimates when breaks are detected.

3. Bubble detection should combine multiple methods (GSADF for statistical significance, LPPLS for timing, sentiment for early warning).

6 Conclusion

This survey has provided a comprehensive review of structural break detection and asset price bubble identification methods, with a novel contribution demonstrating the value of NLP augmentation for early warning systems.

6.1 Summary of Contributions

We have made four primary contributions:

First, we developed a unified taxonomy for structural change detection methods, organizing them by temporal perspective, statistical framework, and output type. This taxonomy enables practitioners to select appropriate methods based on their specific requirements and constraints.

Second, we provided a systematic empirical comparison of detection methods using publicly available data and documented market events. Our analysis shows that PELT achieves the best balance of precision and recall for retrospective analysis, while BOCD provides effective real-time monitoring capability.

Third, we demonstrated that NLP-derived sentiment signals, particularly from FinBERT analysis of financial news, systematically lead structural breaks by 3–7 trading days. This finding suggests that market narratives shift before quantitative detection methods identify regime changes, providing a valuable early warning capability.

Fourth, we released a fully reproducible implementation of all methods in Python, with Docker containerization, continuous integration, and comprehensive documentation. This contribution addresses growing concerns about reproducibility in financial research and provides a foundation for future methodological development.

6.2 Key Findings

Our empirical analysis yields several key findings:

1. Modern changepoint algorithms (PELT, BOCD) outperform classical tests (Chow, CUSUM) in detection accuracy, particularly for multiple break identification.

2. The GSADF test provides reliable bubble detection with a lag of 1–2 months for origination and near-real-time detection of termination.

3. Sentiment deterioration precedes detected structural breaks, with peak predictive power at a 5-day lag.

4. Topic modeling reveals systematic narrative shifts around regime changes, providing contextual information beyond simple sentiment scores.

5. Combined detection approaches achieve modest but consistent improvements over pure quantitative methods, primarily through reduced detection lag.

6.3 Limitations and Future Directions

Several limitations suggest directions for future research:

Multivariate extensions: Our analysis focuses primarily on univariate time series. Extending to multivariate settings would enable detection of breaks in cross-asset correlations and systematic risk factors.

Higher frequency data: The applicability of NLP augmentation to intraday data remains unexplored. Real-time sentiment from social media might provide value for high-frequency applications.

Domain-specific fine-tuning: Pre-trained FinBERT may not capture all nuances of financial language. Domain-specific fine-tuning on labeled break-adjacent text could improve leading indicator quality.

Causal analysis: Our results establish correlation between sentiment and breaks but do not identify causal mechanisms. Experimental or quasi-experimental designs could clarify whether narrative changes cause regime shifts or merely reflect early information processing.

Cross-market analysis: Extending the analysis to multiple asset classes and international markets would test the generalizability of our findings.

6.4 Implications

For researchers, this survey provides a methodological foundation and benchmark results for structural change detection in financial markets. The reproducible implementation enables replication and extension of our findings.

For practitioners, we offer concrete guidance on method selection and demonstrate the value of incorporating textual data into quantitative monitoring systems. The 3–7 day advance warning from sentiment signals, while not sufficient for market timing, provides valuable lead time for risk management adjustments.

For regulators and policymakers, our findings highlight the potential for market monitoring systems that combine traditional surveillance with narrative analysis. Early detection of regime changes could inform policy responses and systemic risk assessment.

6.5 Concluding Remarks

The detection of structural breaks and asset price bubbles remains a fundamental challenge in financial economics. While no single method provides perfect detection, the combination of modern change-point algorithms with NLP-derived sentiment signals offers meaningful improvements over traditional approaches. As textual data becomes increasingly available and NLP methods continue to advance, the integration of narrative analysis with quantitative techniques promises to enhance our understanding of financial market dynamics.

The reproducible framework accompanying this survey provides a platform for continued methodological development. We encourage researchers to extend our analysis to new data sources, asset classes, and detection methods, contributing to the ongoing effort to understand and anticipate structural changes in financial markets.

References

Adams, R. P. and MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B*, 37(2):149–163.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fearnhead, P. and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B*, 69(4):589–605.

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

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

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

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance*, 72(4):1399–1440.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1078.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):324–330.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71(3):421–436.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4):967–1004.

Sornette, D. (2003). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, Princeton, NJ.

Sornette, D. and Johansen, A. (2002). Quantitative analysis of stock price jumps. *Physica A: Statistical Mechanics and its Applications*, 245(3-4):411–422.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3):1139–1168.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

West, K. D. (1987). A specification test for speculative bubbles. *Quarterly Journal of Economics*, 102(3):553–580.

Wilson, A. G. and Adams, R. P. (2010). Bayesian online changepoint detection in multivariate time series. *arXiv preprint arXiv:1010.0802*.

A Computational Details

A.1 Algorithm Pseudocode

A.1.1 PELT Algorithm

Algorithm 1 PELT (Pruned Exact Linear Time)

Require: Time series $y_{1:n}$, cost function C, penalty β
Ensure: Changepoint set τ
$F[0] \leftarrow -\beta$
$R \leftarrow \{0\}$ ▷ Candidate changepoints
for $t = 1$ to n **do**
  $F[t] \leftarrow \min_{s \in R}\{F[s] + C(y_{s+1:t}) + \beta\}$
  $\tau^*[t] \leftarrow \arg\min_{s \in R}\{F[s] + C(y_{s+1:t}) + \beta\}$
  $R \leftarrow \{s \in R : F[s] + C(y_{s+1:t}) \leq F[t]\} \cup \{t\}$ ▷ Pruning
end for
Backtrack from $\tau^*[n]$ to recover changepoint set τ
return τ

A.1.2 BOCD Update

Algorithm 2 BOCD (Bayesian Online Changepoint Detection)

Require: Observation y_t, prior run-length distribution $P(r_{t-1}|y_{1:t-1})$, hazard H
Ensure: Posterior $P(r_t|y_{1:t})$
1: **Growth probabilities:**
2: **for** $r = 0$ to $t-1$ **do**
3: $\quad P_{\text{grow}}(r_t = r+1) \leftarrow P(r_{t-1} = r|y_{1:t-1}) \cdot (1 - H(r)) \cdot P(y_t|r_{t-1} = r)$
4: **end for**
5: **Changepoint probability:**
6: $P_{\text{cp}}(r_t = 0) \leftarrow \sum_{r=0}^{t-1} P(r_{t-1} = r|y_{1:t-1}) \cdot H(r) \cdot P(y_t|r_{t-1} = 0)$
7: **Normalize:**
8: $P(r_t|y_{1:t}) \leftarrow [P_{\text{grow}}, P_{\text{cp}}]/\sum$
9: **return** $P(r_t|y_{1:t})$

A.2 Critical Values

A.2.1 GSADF Critical Values

Critical values for the GSADF test depend on sample size and minimum window fraction r_0. Table 7 provides 95% critical values from 2000 Monte Carlo simulations.

Table 7: GSADF 95% Critical Values

T	100	200	400	800	1600
$r_0 = 0.01 + 1.8/\sqrt{T}$	2.08	2.29	2.46	2.58	2.67

B Reproducibility Details

B.1 Software Environment

```
Python 3.11.0
numpy==1.24.0
pandas==2.0.0
scipy==1.11.0
statsmodels==0.14.0
scikit-learn==1.3.0
ruptures==1.1.7
transformers==4.33.0
torch==2.0.0
bertopic==0.15.0
yfinance==0.2.28
matplotlib==3.7.0
```

B.2 Random Seeds

All random number generators are seeded with value 42:

```
numpy.random.seed(42)
torch.manual_seed(42)
```

B.3 Data Access

All data is retrieved from public APIs:

- Yahoo Finance: No authentication required
- FRED: API key available at https://fred.stlouisfed.org/docs/api/api_key.html
- Finnhub: Free tier at https://finnhub.io

C Additional Figures

See the `charts/` directory for all figure source code and generated PDFs:

1. `01_sp500_structural_breaks/` – S&P 500 with detected breaks
2. `02_method_taxonomy/` – Detection method taxonomy
3. `03_gsadf_bubble_dates/` – GSADF test results
4. `04_lppls_fit/` – LPPLS model fit (Bitcoin 2017)
5. `05_pelt_vs_baiperron/` – Method comparison
6. `06_bayesian_online_realtime/` – BOCD posterior
7. `07_cusum_diagnostic/` – CUSUM test statistic
8. `08_variance_bounds/` – Price vs present value
9. `09_finbert_sentiment_ts/` – FinBERT sentiment time series
10. `10_topic_evolution/` – Topic evolution over time
11. `11_nlp_break_lead/` – Cross-correlation analysis
12. `12_method_performance_heatmap/` – Performance comparison
13. `13_robustness_window/` – Window size sensitivity
14. `14_bubble_detection_timeline/` – Historical bubble timeline
15. `15_conceptual_framework/` – Integrated framework diagram