

Dynamic Early Warning System for Financial Crashes

Mokha Lerthsuwanroj

School of Computing Science

Sir Alwyn Williams Building

University of Glasgow

G12 8RZ

A dissertation presented in part fulfillment of the requirements of the Degree of Master of Science at the University of Glasgow

<Date of submission placed here>

**Abstract**

<Abstract goes here…>

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic form.

<**Please note that you are under no obligation to sign this declaration, but doing so would help future students.>**

Name: Signature:

Acknowledgements

<Acknowledgements go here>

Contents

<Update the table of contents by right-clicking on it and selecting Update Field… and then select page numbers only.>

Chapter 1 Introduction 1

1.1 Motivation 1

1.2 Purpose 1

Chapter 2 Survey 2

2.1 Background Survey 2

2.1.1 Current Methods for Early Warning Systems (EWS) 2

2.1.2 Key Crash Indicator 3

2.1.3 Lags 3

2.1.4 Rationale for Focusing on Index-Level Predictions 4

2.2 Research Objectives 4

2.3 Research Questions 4

Chapter 3 Design and Implementation 6

3.1 Data Collection and Preprocessing 6

3.1.1 Dataset Structure and Columns 6

3.1.2 Labelling Crash Events 6

3.1.3 Handling Class Imbalance 7

3.1.4 Custom Sentiment Scoring using FinBERT 7

3.2 Model Architecture Design 8

3.3 Model Training 8

3.4 Model Evaluation 8

3.5 Out of Sample Analysis 8

Chapter 4 Conclusion 10

Appendix A <Name of appendix> 1

Appendix B <Another appendix> 2

Bibliography 3

# Introduction

## Motivation

Financial market crashes have historically affected to the economic and social disruptions, leading to the downturn, unemployment, and loss of investor confidence. Early detection and intervention are critical for minimizing the impact of such crises on those affected.

In recent years, sentiment analysis has emerged as a powerful tool in financial risk monitoring, offering insights into investor expectations, fears, and behaviours. When sudden shifts in investor sentiment are combined with traditional volatility indicators, they can provide early signals of impending market instability (Liu et al., 2023). However, most existing early warning systems rely on static models that struggle to adapt to the fast-changing dynamics of the modern financial markets (Kustina et al., 2023). By integrating both market and sentiment volatility within a dynamic framework, this research aims to develop more responsive and accurate tool for crash prediction.

## Purpose

The primary purpose of this research is to develop a dynamic early warning system (EWS) that integrates both market-based and sentiment-based volatility indicators to enhance the early detection of financial market crashes. Given the increasing complexity and unpredictability of financial markets, especially during periods of heightened uncertainty, there is a growing need for more adaptive and timely forecasting models.

This study seeks to address by analysing how time-varying patterns in market-based and sentiment-based volatility relate to the occurrence of past financial crises. Specifically, the research will make use of historical financial news data sourced via the Yahoo Finance (yfinance) API to extract relevant sentiment signals. These sentiment indicators will then be combined with traditional market-based volatility measures within a dynamic modelling framework designed to capture the evolving relationships between these variables over time.

# Survey

## Background Survey

### Current Methods for Early Warning Systems (EWS)

Early warning systems (EWS) for financial crashes have evolved from simple statistical models to more advanced machine learning and nonlinear approaches. The traditional statistic models, such as logistic regression, have been widely used to detect early signals for financial crashes using predefined relationships between risk indicators and crash probabilities. Such models often suffer from rigid parameterization and lagging indicators, limiting their ability to capture regime shifts or sudden market changes (Kustina et al., 2023), in contrast with the more recent research that explored the nonlinear approach to overcome this limitation. Nonlinear algorithms, support vector machines (SVM), and neural networks have shown improved capacity for capturing the complex relationships in the real-world financial markets (Song et al., 2024), allowing more flexibility when modelling market risks and crash probabilities as new data becomes available. Empirical evidence supports that dynamic nonlinear methods outperform static models, providing better crisis prediction under changing market environments (Song et al., 2024).

Beside the market-based indicators, sentiment analysis has also gained attention in financial crash predictions. The rise of social media platforms such as Twitter (now known as X), along with financial news sources, has provided rich datasets for capturing investor mood and behaviours (Liu et al., 2023). However, extracting signals from this unstructured data often produces noisy which remains a challenge. Liu, Leu, and Holst (2023) proposed a method using FinBERT combined with an ensemble SVM to reduce noise and filter out irrelevant content from social media discussions.

Huang et al. (2020) showed that FinBERT, which is specifically pre-trained on financial texts including earnings call transcripts, analyst reports, and financial news articles, significantly outperforms general-purpose language models like BERT and traditional approaches in various financial information extraction tasks, including the LM dictionary, NB, SVM, RF, CNN, and LSTM. The model's specialized training on domain-specific vocabulary and financial terminology enables it to better understand the context inherent in financial communications, resulting in improved accuracy for sentiment classification, named entity recognition, and relationship extraction from financial documents (Huang et al., 2020). However, it is important to note that FinBERT demonstrated superiority applies specifically to financial text analysis tasks, and not directly to modelling financial market volatility.

As noted by Parras-Gutiérrez et al. (2014), forecasting models usually designed for short-term or one-step-ahead predictions due to the increasing in difficulty and unreliability of medium- and long-term forecasts caused by error propagation over time. To complement this perspective, Allaj and Sanfelici (2023) introduced a time-varying window (e.g., T = 22, 66, 132 days) in the context of early warning systems for financial instability. This approach acknowledges the changing nature of financial markets and allows models to capture different temporal dynamics ranging within a unified structure. Together, these insights lead to a multi-horizon modelling method that balances predictive accuracy with a greater understanding of time.

### Key Crash Indicator

Most studies operationalize a financial crash using a binary crash indicator equation, where a crash is identified based on a significant drop in asset prices or index returns over a specified time window. A common method involves calculating the log return of closing prices over a fixed period (e.g., 5-day or 10-day intervals) and labelling an observation as a "crash" if the return falls below a predefined threshold which often set at the 10th percentile of historical returns or a fixed percentage drop, such as −10% (Kaminsky et al., 1998). Nonetheless, early warning models also extend to other types of financial crises, such as currency and sovereign debt. Kaminsky and Reinhart (1999), for example, define currency crises based on a sharp depreciation of the exchange rate coupled with reserve losses, using an exchange market pressure index. Bussière and Fratzscher (2006) extend this framework to sovereign debt crises by incorporating a wide range of macroeconomic variables, flagging a crisis when key thresholds are breached.

In addition, volatility also remains one of the most important indicators in crash prediction research. To flag the potential financial instability, both realized volatility (observed historical price variability) and price-volatility feedback rate have been used (Allaj & Sanfelici, 2023). Pattern of increased volatility generally precede market downturns, making it useful for early warning system frameworks.

Mentioning the traditional risk measures, Value-at-Risk (VaR) and Expected Shortfall (ES) are widely used as quantitative measures to assess market risk and potential losses under various conditions. However, both VaR and ES forecasts often rely on models with specific distributional or structural assumptions (Allaj & Sanfelici, 2023), which may not capture sudden market regime shifts, nonlinear behaviours. This is especially true in emerging markets, where volatility is typically higher and market dynamics are less predictable.

A recent study by Le (2024) examined the effectiveness of combining multiple VaR and ES forecasting models in the context of the Vietnamese stock market. The research found that forecast combination techniques, such as weighted averaging of outputs from different models (e.g., GARCH, CAViaR, and ES-CAViaR), significantly improved the accuracy and reliability of risk forecasts, especially during periods of high market volatility. The combined models showed better back testing performance and greater compliance with regulatory risk thresholds, compared to any single model (Le, 2024).

### Lags Selection

In time series analysis, lags involve using past data points to predict the future values. Specifically, a lagged variable is a prior value of the same variable, shifted backward in time by a specific number of time steps. The purpose of including lags is to capture the temporal dependencies, persistence, or self-correlation which commonly found in sequential data such as stock returns, volatility, or macroeconomic indicators (Box, Jenkins, & Reinsel, 2008).

The choice of how many lags to include directly impacts a model’s ability to capture relevant temporal dependencies. Parras-Gutiérrez et al. (2014) addressed this issue in the context of short-, medium-, and long-term time series forecasting using the L-Co-R algorithm, which incorporates a cooperative-competitive evolutionary strategy to automatically select appropriate lags. Their approach revealed that lag structures reflect a broader challenge in time series modelling: too few lags may underfit, missing important dependencies, while too many lags may lead to overfitting or increased computational complexity. The study emphasizes that adaptive or data-driven lag selection methods, such as genetic algorithms or information-theoretic criteria (e.g., AIC, BIC), can enhance model generalizability.

### Rationale for Focusing on Index-Level Predictions

Index-level models offer several advantages, including aggregation benefits that help reduce noise and unexpected shocks from individual stocks. According to Park et al. (2024), the research has shown that top-down index forecasts tend to be more accurate and informative than bottom-up aggregation of individual stock predictions, particularly for systemic risk assessment. By concentrating on index-level sentiment and market volatility, the model can better capture macro-level signals that reflect wide-range market conditions (Park et al., 2024).

## Research Objectives

The goal of this project is to assess the following objectives:

1. To investigate the contribution of both market-based and sentiment-based volatility indicators to financial crash prediction by evaluating their feature importance.
2. To develop a dynamic modelling framework to capture the evolving relationship between sentiment-driven and market-based volatility indicators.
3. To evaluate and compare the predictive performance of the proposed dynamic early warning system against traditional static statistical models, with a particular focus on assessing the added value of sentiment-based inputs.
4. To validate the robustness and generalizability of the developed model across different market environments and historical crisis periods.

## Research Questions

Based on the stated objectives, this study seeks to answer the following key research questions:

1. How do market-based and sentiment-based volatility indicators individually and jointly relate to the timing and occurrence of past financial crashes at the index level?
2. Can a dynamic early warning system outperform traditional static models in predicting financial market crashes?
3. To what extent does sentiment volatility enhance the performance of early warning models compared to using market-based indicators alone?
4. Is the proposed dynamic early warning system model robust and generalizable across different market conditions and historical crisis periods?

# Design and Implementation

## Data Collection and Preprocessing

For this study, the dataset S&P 500 with Financial News Headlines (2008-2024)[[1]](#footnote-1) was utilized, which is publicly available on Kaggle. This dataset combines daily S&P 500 stock market data with corresponding financial news headlines, enabling the analysis of market behaviour alongside sentiment-driven news information. The dataset covers the period from August 2008 through 2024, providing a comprehensive view of the S&P 500's price movements along with market-relevant news during this period.

### Dataset Structure and Columns

The dataset consists of multiple columns, which can be broadly categorized into market data and news headlines:

1. Date: This column records the trading date in the format YYYY-MM-DD. It corresponds to the actual days when the S&P 500 market was open, and trading took place.
2. Headline: This column includes the financial news headline(s) published on the respective trading date. These headlines reflect the key news events, market sentiments, or significant announcements that could potentially influence investor behaviour and market movements.
3. Close: This column provides the closing price of the S&P 500 index on the given trading date, representing the final price at which the index traded on that day.

### Labelling Crash Events

To identify potential future market crashes within the dataset, we introduce a labelling method that flags whether a significant drop in the S&P 500 closing price occurs within a defined future period. This approach allows us to create a binary target variable indicating the presence or absence of a market crash after a given trading day.

The key parameters in this labelling process are:

1. Look-ahead period: The number of trading days into the future over which we examine the price drop. Given approximately 252 trading days in a year, this value is set to 252 to analyse a one-year ahead horizon.
2. Drop threshold: The fractional threshold that defines a crash. If the future closing price drops below this fraction of the current closing price, a crash is labelled. For example, a threshold of 0.9 corresponds to a 10% decline.

### Data Splitting and Class Imbalance Handling

The dataset was divided into training and testing portions using a time-based approach to maintain proper temporal sequence and prevent information leakage. Data collected from 02/01/2008 to 31/12/2021 formed the training dataset, while data from 01/01/2022 to 04/03/2024 constituted the test dataset. However, there is a notable temporal skew in the distribution of crash events within this split. Of the 807 total crash instances, 539 (66.79%) occurred in the pre-2022 training period, while only 268 crashes were recorded in the post-2022 testing period. This imbalance is particularly significant considering that the post-2022 period represents 41.37% of the total data (7,912 out of 19,127 observations). The lower frequency of crash labels in more recent data poses a challenge for time-based evaluation methods, as test sets drawn from the post-2022 period may contain insufficient crash events, potentially leading to undefined performance metrics (e.g., AUC) and biased assessments of model performance. To address this issue and ensure valid evaluation, the train-test split was carefully adjusted to guarantee that the test set contains adequate instances of the minority class.

In addition, to mitigate the highly imbalanced issue with only 4.22% of crash instances, we apply the Synthetic Minority Over-sampling Technique (SMOTE), a widely used resampling method that generates synthetic examples of the minority class based on feature space similarities between existing minority instances (Chawla et al., 2002). By interpolating new samples rather than simply duplicating existing ones, SMOTE improves the generalizability of the model and helps it learn decision boundaries that are more representative of both classes (Budhidharma et al., 2023).

### Custom Sentiment Scoring using FinBERT

To quantify the sentiment expressed in financial news headlines, we employ a custom sentiment score derived from the FinBERT model's output probabilities. FinBERT classifies each input text into three sentiment categories: negative, neutral, and positive, producing corresponding probabilities. Rather than relying solely on discrete class labels, we calculate a continuous sentiment score defined as the difference between the positive and negative probabilities (Hiew et al., 2019):

where are the number of positive, neutral, and negative texts within the period, while are the probability of the headline at time being classified as positive and negative. This formulation captures the net sentiment polarity by balancing positive and negative signals while effectively ignoring the neutral component.

To capture the temporal dynamics and fluctuations in market sentiment, we compute the daily average sentiment score and its rolling volatility over various time horizons. This is formalized as:

where represents the average sentiment score on the day , is the sentiment score of the -th headline, and is the total number of headlines on that day.

Then compute the rolling standard deviation across multiple window lengths, corresponding to different market time frames: one week (5 trading days), one month (22 days), one quarter (66 days), and half a year (126 days). For each window , the sentiment volatility is given by:

where is the mean of over the -day window ending at day .

### VaR and ES Features

To enrich the model with forward-looking risk measures, we incorporate parametric Value-at-Risk (VaR) and Expected Shortfall (ES) as additional features. These measures are computed under the assumption of normally distributed returns using a rolling window approach. For each window length , and standard deviation of daily returns are calculated. The one-day VaR at confidence level is given by:

where is the z-score corresponding to the confidence level (e.g., ≈ 1.64 for 95% confidence). The corresponding Expected Shortfall (ES), which estimates the expected loss conditional on the loss exceeding the VaR threshold, is computed as:

where is the standard normal probability density function evaluated at , and is the tail probability. These risk measures provide a theoretically grounded way to quantify downside risk and help capture volatility dynamics that may precede extreme market events.

## Architecture Design

* Python, framework, library, etc.
* Tw = 5, 22, 66, 126, 252

## Model Training

* Generate sentiment score and compute sentiment volatility
  + FinBERT tokenizer
  + Calculate \*mean\* (group by ‘date’)
* Align with market crashes
* Modelling

## Evaluation Metrics

### AUC

### Confusion Matrix

### Sensitivity

Sensitivity, also known as the true positive rate or recall, is a critical metric in evaluating the performance of early warning systems. High sensitivity is prioritized because it reflects the model’s ability to correctly identify actual crisis events, often the rare but most consequential outcomes. According to Budhidharma et al. (2023), models with higher sensitivity are better suited for early warning purposes, as failing to detect a true crash (false negative) could result in significant financial or systemic losses. Since financial crises are typically low-frequency, high-impact events, sensitivity becomes a more meaningful indicator of predictive usefulness than overall accuracy, which can be misleading in imbalanced datasets. Therefore, in many empirical studies, sensitivity is used as a key ranking criterion when comparing the predictive performance of models for risk and crisis forecasting.

## Out of Sample Analysis

* Test the goodness of the previous model

The content of these chapters depends on the project and should be agreed with your supervisor (e.g. description of the solution, evaluation results, etc).

<Figure below is in style “figure” which continues to style “figure caption” when you press Enter and then back to “Normal” when you press Enter again.>

Figure 1: Some important shapes.

<If you wanted to show any code fragments, you could use the following style called code, which could then be followed by figure caption..>

*# This is a little bit of Python*

**for** i in range( 10 ):

**for** j in range( 10 ):

**print** i\*j,

**print**

Figure 2: A crucial algorithm for the project.

# Conclusion

Main conclusions of your project. Here you should also include suggestions for future work.

###### <Name of appendix>

<Use Heading 6 for the Appendix heading>

###### <Another appendix>

# Bibliography

[1] Kustina, L., Sudarsono, R., & Effendi, N. (2023). Market crash factors and developing an early warning system: Evidence from Asia. *Investment Management and Financial Innovations, 20*(3), 116–125. <https://doi.org/10.21511/imfi.20(3).2023.10>

[2] Song, S., & Li, H. (2024). Early warning signals for stock market crashes: Empirical and analytical insights utilizing nonlinear methods. *EPJ Data Science, 13*, 16. <https://doi.org/10.1140/epjds/s13688-024-00457-2>

[3] Liu, J., Leu, J., & Holst, S. (2023). Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. *PeerJ Computer Science, 9*, e1403. <https://doi.org/10.7717/peerj-cs.1403>

[4] Allaj, E., & Sanfelici, S. (2023). Early warning systems for identifying financial instability. *International Journal of Forecasting, 39*(4), 1777–1803. <https://doi.org/10.1016/j.ijforecast.2022.08.004>

[5] Park, M., Peterson, M., & Weisbrod, E.H. (2024). Top-Down vs. Bottom-Up Index Forecasts: Are Market Strategists Strategically Pessimistic? *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.4695279>

[6] Le, T.H. (2024). Forecasting value-at-risk and expected shortfall in emerging market: does forecast combination help? *Journal of Risk Finance, 25*(1), 160-177. <https://doi.org/10.1108/JRF-06-2023-0137>

[7] Huang, A. H., Wang, H., & Yang, Y. (2020). FinBERT – A large language model for extracting information from financial text. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.3910214>

[8] Kaminsky, G., Lizondo, S., & Reinhart, C. M. (1998). Leading indicators of currency crises. *IMF Staff Papers, 45*(1), 1–48 <https://EconPapers.repec.org/RePEc:pal:imfstp:v:45:y:1998:i:1:p:1-48>

[9] Kaminsky, G., & Reinhart, C. M. (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review, 89*(3), 473–500. <https://doi.org/10.1257/aer.89.3.473>

[10] Jemović, M., & Marinković, S. (2019). Determinants of financial crises—An early warning system based on panel, logit regression. *International Journal of Finance & Economics, 24*(4), 1–15. <https://doi.org/10.1002/ijfe.1779>

[11] Bussière, M., & Fratzscher, M. (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance, 25*(6), 953–973. <https://doi.org/10.1016/j.jimonfin.2006.07.007>.

[12] Parras-Gutiérrez, E., Rivas, V. M., García-Arenas, M., & del Jesús, M. J. (2014). Short-, medium- and long-term forecasting of time series using the L-Co-R algorithm. *Neurocomputing, 128*, 433–446. <https://doi.org/10.1016/j.neucom.2013.08.023>

[13] Budhidharma, Valentino & Sembel, Roy & Hulu, Edison & Ugut, Gracia. (2023). Early warning signs of financial distress using random forest and logit model. *Corporate and Business Strategy Review, 4*, 69-88. <https://doi.org/10.22495/cbsrv4i4art8>

[14] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique*. Journal of Artificial Intelligence Research, 16*, 321–357. <https://doi.org/10.1613/jair.953>

[15] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008). Time Series Analysis: Forecasting and Control. 4th Edition, *Wiley, Oxford*. <https://doi.org/10.1002/9781118619193>

[16] Hiew J. Z. G., Huang X., Mou H., Li D., Wu Q., Xu Y. (2019). BERT-based financial sentiment index and LSTM-based stock return predictability.  *Statistical Finance* *[q-fin.ST]*, ArXiv preprint. arXiv:1906.09024. <https://doi.org/10.48550/arXiv.1906.09024>

1. https://www.kaggle.com/datasets/dyutidasmahaptra/s-and-p-500-with-financial-news-headlines-20082024 [↑](#footnote-ref-1)