Vaasan yliopisto
UNIVERSITY OF VAASA

Elias Keskinen

Predicting a Stock Market Correction

Evidence from the S&P 500 Index

School of Accounting and Finance
Master’s Thesis in Finance
Master’s Degree Programme in Finance

Vaasa 2023

UNIVERSITY OF VAASA
The school of Accounting and Finance

Author:	Elias Keskinen
Title of the Thesis:	Predicting a Stock Market Correction: Evidence from the S&P 500 Index
Degree:	Master of Science in Economics and Business Administration
Programme:	Finance
Supervisor:	Anupam Dutta
Year:	2023 **Pages:** 95

ABSTRACT:
The fundamental pursuit of this thesis is to understand how publicly available information is incorporated into asset prices. The objective is to identify potential inefficiencies within this process and the markets themselves while also meticulously examining how different variables behave in turbulent market conditions. The pace of information absorption impacts all aspects of the financial landscape, influencing academic research, corporate strategies, trading decisions, and regulatory frameworks.

This study utilizes a logistic regression model to forecast imminent market correction movements in the S&P 500 index using a daily dataset that spans from January 2019 to October 2022. A selection of well-researched financial, macroeconomic and option metric indicators serve as independent variables. The research establishes three distinct models, each with a binary dependent variable that designates the occurrence of a crash when market drawdown exceeds either 3%, 4%, or 5%.

These models demonstrate a substantial predictive power for future negative equity returns, predicting negative instances correctly over 77% of the time with in-sample data while statistically significantly rejecting the null hypothesis "predictors at time t have no effect on the likelihood of a crash at time t+1". The significant predictive power is displayed by the Bond-Stock Earnings Yield Differential, Volatility Smirk and Open Interest Difference predictors.

Across the three models, the TED Spread, Bid-Offer Spread, Term Spread, Baltic Dry Index, and S&P GSCI Commodity Index predictors fail to demonstrate consistent statistical significance. Thus, their short-term predictive capabilities should be approached with caution. The out-of-sample test is conducted to affirm the study's robustness, utilizing a training/testing split of 85/15. The best-performing model, which applied a 3% threshold as a binary dependent variable, successfully identified crash observations with 59% precision, thus providing a notable information edge over a random guess. Such findings provide evidence of existing market inefficiencies and the lead-lag relationship between option and equity markets.

KEYWORDS Market crises, stock market corrections, valuation metrics, financial indicators, option metrics, logistic regression, lead-lag relationship, efficient market hypothesis

VAASAN YLIOPISTO
Laskentatoimen ja rahoituksen akateeminen yksikkö

Tekijä:	Elias Keskinen
Tutkielman nimi:	Osakemarkkinan korjausliikkeen ennustaminen: Todisteita S&P 500 Indeksimarkkinalta
Tutkinto:	Kauppatieteiden maisteri
Oppiaine:	Rahoitus
Työn ohjaaja:	Anupam Dutta
Vuosi:	2023 **Sivumäärä:** 95

TIIVISTELMÄ:

Tämän gradun keskeisenä pyrkimyksenä on syventää ymmärrystä siitä, miten julkisesti saatavilla oleva informaatio vaikuttaa osakemarkkinoiden hinnoitteluun. Tutkimus pyrkii paljastamaan mahdollisia epätehokkuuksia informaation hinnoittelussa sekä tarkastelemaan huolellisesti eri muuttujien käyttäytymistä epävakaissa markkinaolosuhteissa. Uuden informaation omaksumisen nopeus liittyy olennaisesti moniin talouden osa-alueisiin, vaikuttaen akateemiseen tutkimukseen, yritysten strategioihin, kaupankäyntipäätöksiin ja markkinoiden sääntelykehyksiin.

Tämä tutkimus käyttää logistista regressiomallia, jonka tavoitteena on ennustaa tulevia markkinakorjausliikkeitä S&P 500-indeksimarkkinassa. Tutkimusaineisto kattaa päivittäisen ajanjakson tammikuusta 2019 lokakuuhun 2022. Aineiston riippumattomia muuttujina toimii erilaisia laajasti tutkittuja talouden sekä optiomarkkinoiden tunnuslukuja tai mittareita. Tutkimuksessa käytetään kolmea eri mallia, joissa binaarinen riippumaton muuttuja osoittaa markkinoiden korjausliikkeen, kun markkinoiden arvonlasku ylittää joko 3 %, 4 % tai 5 %.

Nämä mallit osoittavat merkittävää ennustuskykyä tuleville negatiivisille osaketuotoille, ennustaen negatiiviset tuotot oikein erinomaisella menestyksellä. Mallit hylkäävät tilastollisesti merkittävästi nollahypoteesin "muuttujilla aikaan t ei ole vaikutusta laskun todennäköisyyteen aikaan t+1". Tilastollisesti merkittävimmät muuttujat ovat Bond-Stock Earnings Yield Differential, Volatility smirk sekä Open Interest Difference.

Tutkimuksessa käytetyn kolmen mallin osalta TED Spread, Bid-Offer Spread, Term Spread, Baltic Dry Index ja S&P GSCI Commodity Index -muuttujat eivät osoita johdonmukaista tilastollista merkitsevyyttä. Näin ollen niiden kykyyn indikoida markkinoiden tulevaa arvostustason muutosta tulisi suhtautua kriittisesti. Lopuksi tutkimuksessa testataan mallin kykyä suoriutua oikeissa markkinaolosuhteissa, datalla, jota malli ei ole ennen kohdannut. Menestynein malli onnistui tarkasteluajanjaksolla ennustamaan oikein merkittävän määrän markkinan korjausliikkeistä. Täten, malli ennakoi markkinan suuntaa paremmin kuin sattumanvarainen arvaus. Tämä todistaa, että markkinoilla on yhä epätehokkuuksia, erityisesti sen suhteen, että informaatio hinnoitellaan ensin optiomarkkinoilla ja viiveellä osakemarkkinoilla.

AVAINSANAT: Markkinakriisit, Osakemarkkinoiden korjausliikkeet, Arvostus mittarit, Taloudelliset indikaattorit, Optiomarkkina, Logistinen regressio, Tehokkaiden markkinoiden hypoteesi

Sisällys

Figures

Tables

1 Introduction

The question of how information is reflected in asset prices is fundamental and crucial for everyone involved in financial markets- from academic researchers and corporate managers to traders and regulators. It is a question that impacts the wider economy while also providing remarkable economic incentives for investors. For example, the negative contagion effect from the financial crash of 2008 resulted in millions of people unemployed, and the asset values of pension funds significantly declined. While this was an extreme event, it shows just how devastating socioeconomic impacts may follow market downturns.

Many believe that stock prices reflect all publicly available information, as suggested by Fama in 1970. However, academia and practitioners have increasingly challenged this view, and more recent studies of larger-scale market corrections have often found contradicting evidence (see, e.g., Sornette, 2017, pp 338-346) . This paper is built on the premise that there might be times when stock prices do not fully reflect all publicly available current and past information, especially preceding and during market corrections.

While Sornette (2003) find that market corrections are mainly driven by increased computer trading, asset illiquidity, derivatives usage, deficits, and overvaluation, he posits that the defining factor is the investors' positive herding behavior preceding the crash. Even though he is more concerned with extreme events, his findings and philosophy can be extended to this study, considering smaller scale events. Especially market sentiment is something that is carefully examined. This study not only relies on fundamental or macroeconomic indicators but also tries to take the pulse of market sentiment and the manouvers of so called informed investors via options data.

There is a plethora of studies conducted in an attempt to find factors that best explain the future returns of assets, and it has even produced a phenomenon referred to as "a zoo of new factors," which states that the new variables found are allegedly often guilty of data snooping, and insufficient statistical significance overall (Harvey et al., 2016). As studies are often focused on proving the predictive power of 1 or 2 variables, there is limited research on combining and applying variables showing the most robust results in predicting market corrections. It is a reasonable assumption that combining these variables overall may improve predictions' accuracy. Moreover, studies around this matter often only look for the statistical significance of the variables, but this study aims to see how these variables perform together.

The novelty and significance of this study lie in the unique set of variables used. Each variable is thoroughly evaluated and ultimately tested in a real market environment with previously unseen, out-of-sample data. Furthermore, while many studies around this topic employ complex models, this study utilizes logistic regression, a simple yet robust method that provides easily interpretable results, allowing a wider audience to understand and potentially apply the findings. The dependent variable is also rather unique as it accounts for crash observations from relatively small correction movement, starting from drawdowns of 3%.

1.1 Purpose of the study

The study aims to find a set of variables that could bear information about potential near-term market declines. Utilizing these variables, the goal is to attain over 50% accuracy in predicting these instances of negative equity returns correctly—a level of accuracy that could potentially prove financially rewarding. Thus the intent is to identify the statistical significance of these variables in predicting equity returns and assess their practical performance.

This study defines a stock market correction as a drawdown exceeding 4% from a high watermark, with additional analysis conducted at 3% and 5% thresholds for robustness. The fundamental research question is: Can a correction movement in the S&P500 index at time, *t+1*, be predicted using publicly available data at time *t*? This study is designed to investigate and evaluate the effectiveness of selected variables in predicting such market corrections, with the goal of accepting the alternative hypothesis, $\mathrm{H_a}$.

$\mathrm{H_a}$: $At\ least\ one\ of\ the\ predictors\ at\ time\ 't',have\ an\ effect\ on\ the\ likelihood\ of\ a\ crash\ at\ time\ 't+1'$

1.2 Structure of the study

This study unfolds in six distinct but closely related chapters. Chapter 2 delves into the theoretical concepts the research is based on: the Efficient Market Hypothesis and Behavioral Finance; the former represents one of the foundational concepts in finance, while the latter is a more contemporary field within financial theories. Chapter 3 presents the literature review and discusses previous research papers on predicting (negative) future equity returns. The aim is to find versatile metrics and indicators demonstrating robust predictive capabilities. Consequently, in Chapter 4, the set of variables are selected, introduced and processed, data properties are checked, and model assumptions are validated. An in-depth data-analysis is conducted to fully understand the behavior and scale of the variables and prepare them for the logistic regression model. Chapter 5 presents the empirical results, revealing the outcomes of the study. Additionally, this Chapter compares performance between in-sample and out-of-sample tests, highlighting the effectiveness of the selected predictors. Finally, Chapter 6 concludes the study.

2 Theoretical background

In order to understand financial markets and whether it is possible to detect market crashes or trend reversals, it is meaningful to grasp the underlying theoretical concepts. This Chapter describes these underlying concepts, starting with a review of the efficient market hypothesis (EMH), which, already in its semi-strong form, provides a framework for why financial crashes should not be detectable beforehand without non-public information. Under the assumption of EMH, there should not be methods that allow investors to earn abnormal gains systematically. To address this assumption, this study investigates the market's effectiveness in detecting trend reversals using publicly available information.

The second subsection examines the psychological and emotional factors that influence investor decisions and how these factors may enable the formation of asset bubbles and market inefficiencies. Biases- supported by behavioral finance theories- generally accepted to affect decision-making, such as positive feedback loops, are discussed.

2.1 Efficient market hypothesis

Arguably there is an ongoing disagreement about the efficiency of the equity markets, particularly between practitioners and academics. The question is not whether markets are efficient or inefficient; the disagreement mainly lies in the degree of the market's inefficiency. Practitioners believe there are enough market inefficiencies to obtain excess returns, even after the associated cost of trading. Based on this view, it is reasonable to make efforts and research the markets to obtain those returns. On the other hand, it is often argued that the actions of the market participants result in the elimination of consistent trading opportunities. Only through empirical research can this issue be more deeply understood, and it is desirable that this study also contributes to this question.

Pesaran (2010) summarises that during the early 1970s, economists had widely accepted that stock prices could be accurately represented by a random walk model, meaning that stock returns were unpredictable. In other words, when new information arises, it is rapidly spread and immediately priced in into security prices (Fama 1970: 383-385). Fama was one of the first to make a definitive statement in support of the efficiency of the markets. According to Pesaran, before the development of econometric time series methods and random walk theory, the prevailing theory predicted the movements in the financial markets with the help of business cycles. J. M. Keynes was the most prominent proponent of the relationship between the variation in stock market returns and the stage business cycle.

The EMH was developed in the 1960s from its closely connected theory of the random walk of asset prices put forth by Samuelson (1965). He demonstrated that price changes cannot be forecasted in a market where information is efficiently utilized and thus are unpredictable. In his view, randomness in the stock market results from the many investors eager to increase their wealth and act on even the slightest informational advantage they possess, thus eliminating any existing profit opportunities. In a perfect world where markets are "frictionless," and there is no cost to trading, this process occurs instantaneously, and prices always reflect all available information. Therefore, no profits can be earned through information-based trading because the markets have already priced in the new information. Assets do not have a memory of previous prices, so previous price information is also useless. It may seem counter-intuitive, but Samuelson finds that the more efficient the market becomes, the more random and unpredictable the sequence of price changes generated by the market, and the ultimate efficiency is achieved when price changes are entirely random.

The concept of market efficiency can be assessed using asset pricing models that establish a relationship between risk and asset returns. Among the most renowned of these models is the Capital Asset Pricing Model (CAPM), proposed by Sharpe (1964) and Lintner (1965), which is derived from Markowitz's portfolio theory model and mean-variance efficiency. The model presumes that investors, who are risk-averse and operate with a single-period investment horizon, are primarily concerned with the expected returns and the risk, or variance, of those returns. Hence, investors will always prefer a less risky portfolio for a given expected return, and thus investors willing to take higher risk should also be rewarded with higher returns. Under this model, risk-averse individuals evaluate a speculative opportunity based on its expected return. The model measures the risk associated with an investment using beta. This metric quantifies the covariance of asset returns with market returns relative to the variance of the market itself. Then, as all investors share identical beliefs about the return distributions, they will select an optimal, efficient portfolio to meet their personal risk tolerance.

In the context of the standard expected utility theory, an investor tasked with distributing their wealth between a risk-free and a risky investment will choose to acquire a certain portion of the risky asset if its anticipated future value surpasses its current price (see, Fishburn, 1968). Conversely, if the expected future value of the risky asset falls below its current market price, the investor would opt to short-sell the asset. This behavior encapsulates the core principle of maximizing utility in decision-making.

Pesaran (2010) argues that although some studies found statistical evidence against the random walk hypothesis, a close relative of EMH, they were economically unimportant and statistically suspect. It is known that market efficiency can persist even when market participants hold differing beliefs and act irrationally. This is possible if individual mistakes are not heavily related to each other, a state known as weak cross-sectional dependence. However, during periods of market euphoria, bubbles and crashes can occur. This is when individual errors become strongly interdependent, or in other words when individual mistakes are heavily linked. These periods of strong cross-sectional dependence can often mark deviations from market efficiency.

Fama (1970) distinguishes three different degrees of market efficiency: weak, semi-strong, and strong forms. The three forms of the EMH vary in how well public and private information is incorporated into the asset prices. The weak form asserts that all price information is fully reflected in asset prices, implying that future prices cannot be predicted from past price movements. The semi-strong form requires asset price changes to fully reflect all publicly available information. This implies that no one can achieve superior returns by trading on publicly available information, including financial statements, macroeconomic news, or past price movement.

The strong form states that market prices reflect all information, even information that some market participants have exclusive access to, also known as insider information. Fama (1970) finds the strong form useful as a benchmark for evaluating other forms of market efficiency, not as a form that should be supported. He concluded that the empirical evidence supported the weak form, while the evidence for the semi-strong form was limited.

Fundamental analysis aims to gain the informational edge by understanding the firm's intrinsic value better than others. Hence, it is pertinent for a superior analyst to understand the major variables relevant to the valuation process. Supporting semi-strong form, the EMH posits that fundamental analysis is bound to fail in beating the market (Fama, 1970). The argument is that analysts' findings based on publicly available earnings and other company information are not likely to be more accurate or insightful than competitor analysts’.

In the case of EMH supporting the weak form, technical analysis is invalid because by the time information is public, the price adjustment would have already taken place, and technical trading systems based on past trading data would have no value. In contrast, technical analysts believe that investors take time to analyze information leading to slower price discovery, thus giving an edge to the technician. An example is given by Lo et al. (2000), who find that technical patterns give traders an informational edge and that using algorithms could further increase its applicability. However, they note that this information edge is not always enough to result in excessive profits.

However, it is helpful to separate concepts of efficiency and predictability, as Pesaran (2010) provides a mathematical framework to prove that stock market returns will be non-predictable only if the market efficiency also fulfills the condition of risk neutrality of investors, which arguably is not the case in the financial markets. Further, Fama and French (1988) noted that the predictability of returns in the context of intertemporal or time series models does not necessarily conflict with the concept of market efficiency. Thus, it can be concluded that there are predictable components in stock market prices. The main debate is about the economic factors that drive predictability in asset returns, which relates to controversies around the efficient markets hypothesis, stock market rationality, and the existence of abnormally profitable trading strategies (Lo and MacKinlay, (1999), p.283-284).

Malkiel (2003) summarises that critics of market efficiency argue that there are recent instances where market prices cannot be attributed to rational investors, and psychological factors play a dominant role. The behavior of irrational investors, rather than rational ones, was responsible for setting market prices. For example, from early to mid-October 1987, the stock market lost a third of its value without any significant change in the general economic environment. This suggests that market prices could not have been efficient both at the beginning of October and in the middle of the month. Similarly, it is widely believed that the pricing of Internet stocks in early 2000 was based on the behavior of irrational investors.

2.2 Behavioral finance

This section aims to provide an understanding of behavioral finance, a more recent area of study in financial theories that attempts to enhance our understanding of the financial markets by collaborating with other social sciences. Unlike traditional theories that assume investors to be rational, behavioral finance acknowledges the possibility that investors may act irrationally. As behavioral finance literature is rather extensive, this section concentrates particularly on the psychology and human behavior contributing to stock market inefficiencies. A few central themes are considered to understand better why asset price inefficiencies may form, and whether these phenomena are detectable.

The efficient market hypothesis has been challenged increasingly, and research has shown that the theory of random walk tend to be violated by the asset price fluctuations (see, e.g., Fama 1995, Lo and MacKinlay 1988, and Shiller, 2003). Shiller (2003) specifically underlines the innovative contribution of behavioral finance, which enriches financial research by incorporating factors that conventional models fail to accommodate. This approach broadens our understanding of financial markets by considering various cognitive, social, and psychological aspects, contributing to a more holistic view of investor behavior and market dynamics.

Ritter (2003) outlines that behavioral finance revolves around two main concepts. The first is cognitive psychology, which looks into how individuals involved in the markets make decisions and process new information. The second concept is the "limits to arbitrage," which refers to the circumstances that could prevent market participants from taking advantage of market inefficiencies. In simple terms, it acknowledges that, in real-world scenarios, not all market inefficiencies can be easily exploited for profit due to various factors, like transaction costs or systemic risks. Psychological research shows that individuals tend to make predictable errors in reasoning: they frequently exhibit overconfidence and place undue emphasis on recent experiences, among other things. Behavioral finance harnesses this knowledge base instead of disregarding it.

Speculative trends tend to override fundamental beliefs. As Sornette (2003) simplifies that investors' decisions to buy or sell stocks are responsible for the movements of stock prices. Any deviation from a random walk must ultimately be attributed to their behavior.

Positive feedback loops are a central topic to this hypothesis, where the observation of recent high market returns makes it more likely that price increases will continue, and vice versa in market decrease. This concept bears similarities to the economic principle of increasing returns, where economies of scale allow for a lower cost per unit as the quantity of output increases. Positive feedback loops can lead to a situation where prices diverge too much from the equilibrium, creating a bubble and subsequent crash or correction movement.

The stock market and investor behavior encompass many mechanisms that may generate positive feedback. One widely recognized mechanism, known as the "herd" or "crowd" effect, is based on the imitation behavior of investors (Sornette, 2003). For instance, a group of investors mimicking the actions of others can lead to collective decision-making that is based on limited information, potentially neglecting essential factors such as current news or financial reports.

Herding behavior in speculative markets has been gaining empirical support, as evidenced by numerous studies (see, Shiller, 2000). Sornette (2003) summarises that herding has been linked to various economic activities, such as investment recommendations, price behavior of initial public offerings, earnings forecasts, corporate conservatism, and portfolio management. Herding behavior can be observed as an information cascade in which every subsequent actor makes the same choice based on the observations of the actions of others, irrespective of their knowledge.

Shiller (2003) argues that while EMH insists that when irrational optimists buy more equities as a result of a positive feedback loop, so-called informed investors or “smart money” should offset the effect of these positive feedback loop traders, the smart money does not succeed in offsetting the impact. As such, the dilemma of whether smart money can counterbalance market prices back to fundamental values is widely argued. For example, De Long et al. (1990) find that smart money may even do the exact opposite: they might front-run the feedback traders by buying the security before them in anticipation of further price increase. This behavior drives the prices further away from the fundamentals.

Miller (1977) theorizes that when smart money supposedly knows that security is overvalued, they might still refrain from doing so as they find that short-selling costs are too high. This also creates a reason for this study to use option data, as options provide an extremely efficient way to express interest in profiting from the market's decline. Also, as it is in options nature, the buyer of the option has limited liability with leverage, which may provide further incentive to express opinions through option markets, increasing the total volume simultaneously.

Lyócsa et al. (2022) give a recent example of herding behavior in "Yolo trading", where distinct features of taking an investment action were based on rather vague recommendations of other anonymous users of Reddit's social media platform. More specifically, they study retail investors' decentralized short squeeze strategies in 2021 and 2022 on certain publicly traded companies. These strategies are known to lead to inflated prices or bubbles and, as such, are highly risky. Analysis of these events indicates that the discussion on the Reddit platform intensified in tandem with price variations in the target stocks. Although not all of the price variation was directly linked to the discussions, they conclude that such social network activity has the potential to activate other retail investors to participate in short-squeezing institutional investors, and by extension, such trend could potentially lead to stronger herding behavior.

Building on the evidence from Lyócsa et al. (2022), social media platforms like Reddit or Youtube become powerful accelerators of behavioral finance phenomena. With its emphasis on community-driven sentiment and peer-to-peer influence, social media amplifies behavioral biases such as herd mentality, FOMO (fear of missing out), and overconfidence, among others. Traders, especially novices, may rely heavily on social cues and collective wisdom in their decision-making, thereby fueling market volatility and potential inefficiencies. These digital platforms provide a stage for campaigning various investment strategies and magnify the reach and impact of such behavior.

It is worth noting that herding behavior is not solely a retail investor phenomenon. Even professional investors, such as institutional fund managers, exhibit herding tendencies. This can be attributed to performance anxiety, reputation management, and industry norms. Such behavior can have far-reaching effects. For instance, institutional herding can trigger a chain reaction, leading to inflated prices or market bubbles. These trends can create market distortions and potential long-term financial instability, underscoring the need for continual vigilance and understanding of herding behavior within our increasingly digitized financial markets.

Finally, Ritter (2003) notes that although behavioral finance often supports that the markets are informationally inefficient, it does not mean this inefficiency is easily exploited for profit. For instance, he points out that investors who short-sold outrageously overvalued Japanese stocks in 1988 or the U.S. technology, media, and telecom (TMT) stocks in 1999 faced substantial financial losses. These losses occurred as the overvalued stocks continued to move on their upward trajectory. Eventually, the overvaluation started to unravel, but at this point, short sellers were already financially incapacitated and had no means to recover their losses. As such, Ritter argues that the dynamics of arbitrage function less effectively in the context of low-frequency events. However, interestingly for this study, he suggests that they operate much more effectively in high-frequency situations.

3 Literature review

This literature review focuses on identifying a strategy and, most importantly, variables with academic support for predicting market downturns. While academics widely considered that prices follow a random walk until the 1970s, more recent empirical research has challenged this and has provided evidence that equities exhibit some degree of predictability.

This predictability can either stem from the stock returns' historical data or from other publicly available information, such as financial reports and macroeconomic news. These predictive indicators could include factors like option volatility, price-earnings ratios, short and long interest rate changes, fluctuations in industrial production, or even calendar-based patterns. Nevertheless, the research about indicators should be approached and analyzed carefully as data-snooping bias is involved, thus possibly facilitating spurious regressions (Sullivan et al. (1999)). Data snooping is a term used when the same set of data is used multiple times for making decisions or choosing a model. This could lead to patterns being found that are not really there; they just happened to find positive results by chance in that specific dataset.

Harvey et al. (2016) expand the notion of Sullivan et al. (1999) and argue that as dozens of new indicators of future equity returns are seemingly found on a continuous basis, a much higher hurdle rate is needed to deem factor statistically significant. They propose a threshold t-statistic of 3, instead of the norm of t-statistic threshold of 2. Out of 296 variables tested, they deem approximately half of them false discoveries, meaning that these do not predict equity returns. Their discovery flags the importance of careful due diligence when selecting and analysing the effectiveness of the variables used.

While this paper does not attempt to predict equity returns per se, it does attempt to predict market corrections, measured as drawdowns. Some of the most important differences are that when it comes to market correction movements, there may often be triggers or other warning signs that precede them. For example, these could be factors like overly optimistic investor sentiment or high valuation levels.

A drawdown is defined as a continuous decrease in the asset's price that might be happening over multiple days. It is the cumulative loss from the last maximum to the next minimum price. So essentially, drawdown is the statistic that displays the significance of the market crash or the cumulative loss the investment suffers (Sornette, 2017, pp 51-52). While the drawdowns might fluctuate in their significance, the ability to predict such events would greatly increase long-term investors' risk-adjusted returns (Berge et al., 2008). Further, as S&P 500 index is a prominent leading indicator of economic growth, economists and governments should also have a profound interest in predicting drawdowns (Hertzberg and Beckman, 1989).

A stock market crash is an event of often rapid and unanticipated drawdown, or in other words, a depreciation across a significant cross-section of the stock market. Among academia, there is no consensus on the steepness of market decline used to define market correction or bear market, as discussed by Candelon et al. (2008). However, the smaller the drawdown is, the harder it might be to predict, and vice versa; the larger the drawdown is, the more likely some symptoms are detectable in the economy preceding the event. Chen (2022) notes that buying stocks on a margin before this sudden drop and selling shares once it has happened are the two most common ways investors can lose money when the market crashes.

3.1 Predicting stock market declines

Predicting stock market declines and movements all together has long been a focal point of financial research and has inspired the development of various theories and models. This Chapter reviews the existing literature on methods and indicators used to predict market downturns. An attempt is made to research option metrics along with financial and macroeconomic indicators with leading capabilities relative to the stock market. As predicting the stock market is notoriously difficult, many variables are researched and, consequently, used in this paper.

3.1.1 Option market metrics

This Chapter aims to provide recent literature surrounding Black's (1975) finding that information is first revealed in the options market. As Easley et al. (1998) conclude, in a market with frictions, such as short sale restrictions, informed investors prefer to hold options rather than equities mostly due to embedded leverage. One of the most commonly used metrics for estimating the actions of informed traders is volatility skew, also known as volatility smirk.

Xing et al. (2010) investigate the predictive power of volatility smirks and the delay in which the underlying incorporates the information embedded in the volatility smirk. By using a sample of S&P 500 companies, they find that the volatility smirk is a strong predictor of future negative returns of the underlying. The authors define the smirk as the difference between the implied volatilities of out-of-the-money (OTM) put options and at-the-money call (ATM) options.

Xing et al. (2010) document that the economic reasoning behind the variable is that informed traders may choose to trade in certain markets first, and that market is the one that the information is first embedded in the prices. For example, institutions might prefer to trade out-of-the-money put options and use them as insurance. They argue that this creates so-called lead-lag relationship with the options market and equity market as they find that implied volatilities of OTM put options and ATM call options provide both, predictability and information content about future equity returns.

The volatility smirk pattern is a well-known phenomenon in finance. For example, Bates (1991) gives a comprehensive overview of the stock market crash of 1987 and shows strong evidence of how OTM puts became unusually expensive preceding the crash. His study was made with index options, which indicate that smirk also reveals systematic risk and not only stock-specific risk. Supporting this argument, Pan (2002) divides the components of the option pricing model into jump risk premium and volatility premium and finds that 80% of the OTM put option premium is derived from jump risk premium. This is a significant indicator that informed investors are precisely using OTM puts to express their views on possible negative price jumps, which causes OTM put prices to rise.

Using this literature as a background for their paper, Xing. et al. (2010), find that there is a persistent pattern when applying these principles into individual stocks during 1996-2005. On a risk-adjusted basis, stocks with a steeper volatility smirk underperform those stocks with a flatter smirk by 10.90% per annum. Further, they continue to research how fast the information from volatility smirks is incorporated into equity prices. They discover that the volatility skew's ability to predict future stock returns remains from immediate future up to six months. Their final discovery is that the stocks with the worst quarterly earnings shocks are those with the greatest volatility smirks. This indicates that the volatility smirk includes enough information to predict negative earning surprises to some extent. This finding suggests that volatility smirk is connected to business fundamentals.

There are also some opposing views. Heston (1993) shows that option pricing models are independent of the probability distribution of the underlying asset returns. For example, the well-known Black-Scholes model is independent of the expected stock return. Consequently, he develops a stochastic volatility model that assumes perfect information flow between option and stock markets. The model produces a volatility smirk, but it has no predicting power of underlying stock returns as the expected return is not a variable in option pricing. So this would suggest that, in theory, volatility smirk would not provide reliable predictions on future returns. However, in practice, there is not a perfect information flow between the stock and option markets, as Xing et al. (2010) argue.

Easly et al. (1998) provide evidence that stock returns could be predicted with only seller- and buyer-initiated option volume, while Ofek et al. (2004) find evidence using the no-arbitrage principle and put-call-parity. They document a strong relationship between rebate rate spread, used as a measure for short-selling restrictions, and the magnitude of put-call parity violations. However, applying their findings, the short-selling constraints are gradually diminishing, so the relationship should hardly be arbitrageable. Nevertheless, their findings support the behavioral basis of this paper.

Similarly, Fodor et al. (2011) are studying the information flow between equity and option markets. Their research adds robustness to previous studies that a lead-lag relationship exists between option and equity markets. The study's main finding is that open interest difference, defined as the amount of outstanding call option contracts subtracted from the amount of put option contracts, is a strong predictor of future negative returns of the underlying. Open interest refers to outstanding option contracts that have not been closed, expired, or exercised. Firms with recent negative open interest difference change will significantly underperform those with positive change.

Correspondingly, Gârleanu et al. (2009) argue that the open interest difference measure does not merely mirror current market conditions. Instead, it also provides valuable insights about the likelihood of extreme returns in the future. This stems from the fact that even if it is theoretically possible, market-makers and proprietary traders cannot perfectly hedge their positions. This leads to potential risks and imbalances in the market, which can be reflected in the open interest difference. Therefore, a substantial change in the open interest difference can signal that market participants anticipate price jumps in the underlying.

Glosten and Milgrom (1985) argue on behalf of bid-ask spread and its role before volatility or information events. They conclude that the bid-ask spread widens when traders with superior information edge to market makers enter trades, thus predicting market movements. Similarly, in an attempt to explain the wide spreads of the option markets, Kaeck et al. (2022) confirm that in liquid option markets where the underlying is S&P 500, the spread mainly stems from asymmetric information in the markets and market-maker inventory effects.

Similar findings were made by Choi and Jayaraman (2009) in their event study of stock market overreactions. They document that the bid-ask spread is a strong part of the price discovery process, especially before large price declines. The option market bid-ask spread is more sensitive than the benchmark spread, leading to possible signals on forthcoming changes in the stock market. For the context of this study, they find that the percentage spread increases for both call and put options relative to their benchmark spread before the large price decline date, but after the event, the spreads often quickly decline.

3.1.2 Financial indicators

Financial metrics provide a quantifiable measure of a company's performance, financial health and growth prospects. These metrics are often derived from companies' financial statements and are widely used among investors and fundamental analysts to evaluate companies and the condition of markets overall. In the context of crash prediction, financial metrics can offer valuable insights into market trends and economic conditions that could lead to a market correction.

Certain financial metrics can be particularly informative. For example, price-to-earnings ratios can provide insight into whether the market is overvalued, as historically high price-to-earnings values often precede market corrections (Shiller, 2015). Similarly, Lewellen (2004), focusing on short-term returns, confirms the predictive power of such variables as dividend yield, earnings yield, and book-to-market ratios.

Berge et al. (2008) suggest that a framework provided by the Federal Reserve, called the Fed model, is superior to price-to-earnings ratios in its predictive power. The model takes the difference between equity earnings yield and Treasury bonds. The model is based on the notion of optimal asset allocation between equities and bonds. When bond yields are high relative to equity yields, the model suggests that an adjustment is needed, and as a result, money flows from equities to bonds, causing a price correction.

Following the framework, Lleo and Ziemba (2017) test the predictive power of the Bond-stock earning yield differential (BSEYD) and find significant results. BSEYD model uses daily price, monthly earnings, and prevailing 10-year or 3-month Treasury yield. It is a close relative of the Fed model, with the main difference being that the fed model assumes that earnings yield is equal to bond yields in the optimal state. Instead of asserting this kind of equilibrium state, BSEYD measures the differential between the bond yield and the earnings yield. This differential indicates how far the market is deviating from the so-called equilibrium condition suggested by the Fed Model.

The unique aspect of the BSEYD model is that it is designed not to predict specific performance metrics but to anticipate rare market events such as stock market crashes. This makes it particularly suitable for this study, which focuses on optimizing the likelihood of such crashes impending. Maio (2013) reinforced this view by demonstrating that the 'yield gap', the difference between the dividend or earnings yield and long-term Treasury yields, can offer better results than models based solely on dividend yields. Not only does this approach provide more accurate predictions, but it also aligns better with economic logic. It gives a more comprehensive understanding of why investors might opt to increase their bond holdings and reduce their equity allocation. Investors can make more informed decisions based on potential earnings by considering the yield difference rather than just the stock price or dividend yield separately.

Examining other metrics, Boudt et al. (2017) note that TED spread, defined as the difference between a 3-month LIBOR (USD) and a 3-month U.S. Treasury bill, provides valuable insight into equity markets' liquidity regimes. Similarly, Cheung et al. (2010) highlight the importance of TED spread in the context of the Global Financial Crisis of 2007-2009. More specifically, during a crisis, unexpected changes (shocks) in the U.S. market have at least twice the usual effect on other global markets. However, the impact of shocks from the TED spread on global market indices increases by at least five times. In other words, during a crisis, changes in the perceived credit risk in the economy have an even more significant effect on global markets than shocks from the U.S. market. Thus the TED spread serves as a leading fear indicator in the markets worldwide, including the U.S.

3.1.3 Macroeconomic indicators

Several studies have highlighted the role of different macroeconomic variables in forecasting market corrections and their potential impacts on equity markets. Thus, an exploration of market crash predictors would be incomplete without an analysis of macroeconomic variables, of which inflation is a notable catalyst for market corrections. As Modigliani and Cohn (1979) detailed in their study, levered firms operating in a market influenced by inflation can see a long-lasting impact on their reported earnings, even leading a profitable company towards increasing losses. They argue that in order to maintain the same level of real debt, a firm must increase its nominal debt at the rate of inflation, leading to an increased amount of funds facilitated to the service of the debt. This perspective furthers the argument of market inefficiencies, as the researchers provide evidence of significant misvaluation in the S&P 500 at the end of 1977. Given that the 1970s was marked as "the decade of inflation", these insights could bear pertinent implications for our current high-inflation economy, shedding light on potential areas of concern.

In his paper, Chen (2009) researched numerous macroeconomic variables and whether they have predictive power over equity bear markets. Based on S&P 500, a robust analysis concludes that term spread and inflation are the most statistically significant and economically meaningful macroeconomic indicators. He adds that predicting recession and bear markets with macroeconomic variables is much more successful than attempting to predict equity returns continuously.

Expectations about future economic scenarios that could impact the stock market are often associated with macroeconomic variables. The factors such as unemployment rates, inflation rates, and economic growth could influence future investment and consumption behaviors, thus affecting money flow into or out of equities. Correspondingly, Rapach et al. (2005) demonstrated in their multi-country study on industrialized nations that interest rates emerge as the most robust macroeconomic indicator for predicting equity returns.

Hjalmarsson (2010) presents evidence that term spread, defined as the log difference of 10-year Treasury bonds and 3 month Treasury bills, is a significant predictor of stock returns in developed markets. The relationship between term spread and equity returns is strong and positive. He also finds some contrasting evidence of earnings yield and dividend yield ratios not being robust in predicting power. Thus he recommends focusing on interest rate-based variables in analysis to predict the stock market. Estrella and Trubin (2006) center their research more strongly on the duration of Treasury instruments used and find that most informational spread is between 10-year and 3-month rates. They also find more evidence in favor of the level of the term spread being the significant factor in the variable rather than change in it. They conclude that the flattening of the curve indicates a decrease in economic activity, and the successfulness of the variable is based on the accurate depiction of future economic activity.

As previously mentioned, inflation stands out as one of the most robust predictors of equity returns and market corrections. Mensi et al. (2013) provide insights into the volatility spillover between commodity and equity markets. Specifically, they investigate the predictive power of multiple commodities, such as wheat, oil, and gold, on the S&P 500 returns and, reciprocally, the influence of the S&P 500 on these commodities. The phenomenon studied, known as volatility spillover, is characterized by fluctuations within one financial market or asset class that subsequently impact another. Their findings underscore the significant correlation between oil volatility and the volatility of the S&P 500, particularly during periods of economic uncertainty or financial instability. The findings are bi-directional in the sense that S&P 500 returns are a strong predictor of commodity returns as well.

Correspondingly, the principal component and factor analysis of Liang et al. (2020) reveals that grains and softs, especially, have a significant predictive power of future realized volatility of the S&P 500. Similarly, they underscore that the predictive utility of these relationships is particularly strong during periods of economic downturns and preceding crises. The volatility transmission from these commodities to the stock market becomes particularly potent during these turbulent times, making them invaluable in predicting and navigating market uncertainties.

4 Data and methodology

The data set used in this paper is sourced from 1) DeltaNeutral, which provided daily S&P 500 index European put and call option prices, implied volatilities, and option Greeks, from the Chicago Board Options Exchange (CBOE), and 2) Refinitiv, which supplied the remaining daily financial and macroeconomic data. This comprehensive dataset covers all strike prices and maturities, excluding options with zero volume or zero open interest, to ensure the data's relevance and accuracy. Alongside the options data, relevant financial and macroeconomic data for the same period is continuously published by Refinitiv.

This Chapter introduces the specific definitions of market crashes used as the dependent variable in this study, followed by an explanation of how these observations are gathered. Subsequently, the independent variables, which later form the basis of the logistic regression model, are scrutinized, and mathematical frameworks used to modify these variables are introduced. Additionally, a visual inspection and analysis are conducted to examine whether these variables could have the possibility to provide enough information to enhance the model's edge. Thereafter, the descriptive properties of the processed variables are examined.

The methodology, the logistic regression framework, is introduced after this examination. Lastly, the crucial assumptions of the logistic regression model are discussed, and the data is further analyzed. Conclusively, this Chapter verifies that the data fulfills the necessary conditions and is ready for the logistic regression analysis.

4.1 The crash definition

As a base scenario, this paper identifies a market crash when S&P 500 declines by at least 4%. In addition, for further robustness, crash thresholds of 3% and 5% are examined. Most of the figures relating to the two other thresholds are located in the Appendix for better readability of the paper. Peaks and troughs are identified from the time series, and market decline observations meeting the drawdown criterion are classified as crashes. In other words, every trading day between high and low points is considered a single crash observation. See Figure 1 for a more detailed illustration of crashes of 4% or more. Periods of market decline are marked with vertical red areas.

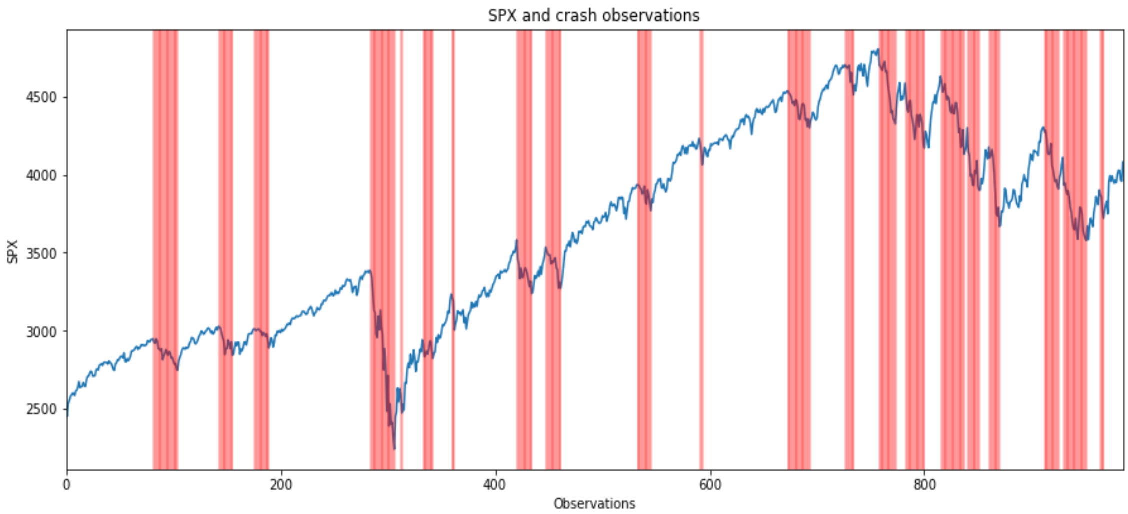


Figure 1. S&P 500 and crash observations of 4% or more

Table 1 provides further insights into crash statistics, including drawdown data. By examining the historical data of drawdowns and durations, the nature of the crashes can be better understood. The study period of this paper is four years (2019-2022); during this time, the criterion introduced above delivers a total of 21 crashes, including 278 crash observations out of 964 observations.

The average drawdown in the dataset varies between 7.47% and 10.39%, depending on the threshold used. These, on average, almost 2-digit drawdowns indicate a strong economic incentive for detecting crashes. The average crash duration is 19 calendar days, not trading days, which means there is also a wide angle of opportunity to act to this decline. The largest crash corresponds to the COVID-19-related stock market decline, which resulted in a 33.92% market decline. Duration time also differs, with the shortest market decline only lasting for two days and the longest lasting 34 days.

For the logistic regression model, the balance of the observations of the dependent variable is central. Naturally, the best balance is at the 3% threshold, with a total of (309/964) 32% crash observations in the dataset. At the 5% threshold, the balance declines to only 23.4%. Additionally, for the most stringent threshold, there are only 15 crash events; for the most lenient definition, there are 27 crash events during the period.

$$Drawdown = (SPX_{peak} - SPX_{through}) / SPX_{peak}, \tag{1}$$

Table 1. Drawdown summary

	Drawdown Threshold		
	3%	4%	5%
Number of crash events	27	21	15
Total crash observations	309	278	226
Average drawdown(%)	7.47	8.61	10.39
Average duration of the crash (days)	16	19	21

Figure 2 introduces the distribution of the drawdowns. As expected, most drawdowns are in the 4 to 10 percent range. There were a total of five 2-digit drawdowns, which is typically used as a threshold for the market decline to be called a "bear market" or "crash". However, academia and practitioners have no clear consensus on the bear market definition. Notably, the distribution of the duration of market crashes is fairly spread out between the minimum and maximum duration. See a more detailed table in the appendix.

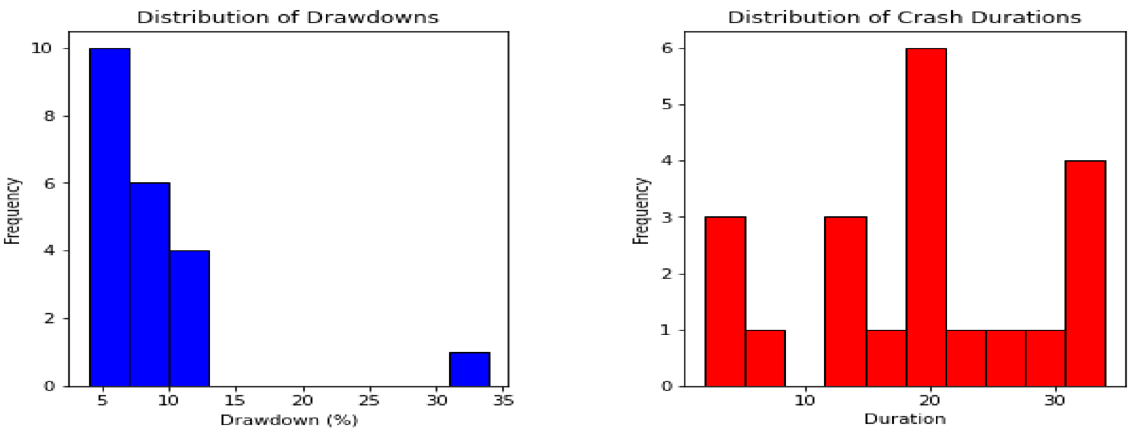


Figure 2. Distribution of drawdowns, 4% threshold

4.2 Data and Analysis of S&P 500 Index Options and Financial Indicators

This section contains insight into how variables are mathematically formulated and a deeper analysis of the variables used. Additionally, the section introduces the processes undertaken to modify the variables to achieve stationarity, follow other researchers' work, or fulfil conditions required by the logistic regression model. A total of 8 variables are introduced and examined in depth. The basis for selecting the variables was introduced in the literature review.

Initially, three option metrics that attempt to provide information on the future of the underlying are introduced: volatility smirk, open interest difference, and bid-offer spread. Subsequently, five financial and macroeconomic metrics are analyzed: Bond-Stock Earnings Yield Differential, TED-spread, S&P GSCI Commodity Index, Baltic Dry Index, and Yield Curve. These metrics play a crucial role in capturing the broader economic conditions that may ultimately impact the equity indices.

Table 2 provides an overview of the variables in their unprocessed form before any transformations or adjustments are made. The table is then referred through this section when discussing the variables. Certain variables, namely skew, spread, TED spread, BSEYD, and the yield curve, are directly introduced as percentage values in the table to improve interpretability. The rest variables are given as their actual values.

Table 2. Summary statistics of the unprocessed variables

	Mean	SD	Range	Min	Max	Kurt	Skew
Skew (%)	4.00	2.88	22.58	- 1.91	20.67	4.96	1.96
OIX	252.21	4.47	10 778.91	- 5 704.06	5 074.85	0.70	- 0.37
Spread (%)	1.89	1.09	10.05	0.58	10.63	12.93	2.79
TED_Spread (%)	0.26	0.20	1.44	- 0.02	1.42	9.84	2.66
BSEYD (%)	3.20	0.69	5.70	0.53	6.23	1.72	- 0.14
Yld_curve (%)	1.19	0.79	2.90	- 0.36	2.54	- 1.34	0.02
S&P_GSCI	2 578.64	68 964.48	3 045.41	1 249.25	4 294.66	- 0.39	0.51
BDI	1 829.26	99 391.00	5 257.00	393.00	5 650.00	1.36	1.07

4.2.1 Volatility smirk

To follow the methodology of Xing. et al. (2010), options data is first filtered to include only options with days till expiration (DTE) anywhere between 5 and 60 for volatility smirk calculation. Because this study is similarly focusing on short-term fluctuations, options not meeting this criterion are not considered.

The moneyness of the S&P 500 index options is calculated as Strike, *K,* of an S&P 500 option, *i,* divided by S&P 500 index, *i*, at a price, *S*, at a matching date, *t*.

$$Moneyness = \frac{K_{i,t}}{S_{i,t}} \quad (2)$$

The average implied volatility (IV) of the at-the-money (ATM) call option is derived from all options with moneyness within the range of 0.95 to 1.05. Respectively, the average implied volatility of the out-of-the-money (OTM) put option is filtered from options within the moneyness range of 0.8 to 0.95. Once grouped, Table 3 results reveal that, as expected, put options are more expensive and volatile on average. Furthermore, the put option distribution exhibits marginally higher skewness and kurtosis than its call option counterpart. Both distributions, however, demonstrate right-skewness, implying that the observations have a long right tail and contain some extreme values.

Table 3. Implied volatilities

	mean	std	min	max	skewness	kurtosis
SPX	3 619.49	639.53	2 237.40	4 807.39	0.06	- 1.30
Call_IV (%)	16.15	7.27	0.00	67.15	2.10	8.24
Put_IV (%)	20.15	8.67	0.00	85.26	2.27	9.71

Following Xing et al. (2010), this study calculates implied volatility smirk, $SKEW_{i,t}$, for S&P 500 index, i, options at day t, as the difference between the implied volatilities of OTM puts and ATM calls, indicated as $VOL_{i,t}^{OTMP}$ and $VOL_{i,t}^{ATMC}$.

$$SKEW_{i,t} \ = \ VOL_{i,t}^{OTMP} - VOL_{i,t}^{ATMC}, \ t = 5, \ldots, 60 \ , \tag{3}$$

Skew (%) has a mean of 4.00% and ranges from -1.91% to 20.67%, with a positive skewness of 4.96. Thus, on average, OTM put options have been 4% more expensive than ATM call options. In extreme instances, puts have been less expensive than calls, indicating strong bullish or optimistic sentiment. On the other end of the spectrum, puts have been as much as 20.67% more expensive, which signals strong bearish sentiment across the market.

Examining Figure 3 below, from the perspective of trying to understand if volatility smirk could be a potential variable for estimating market crashes, it is evident that, for example, before and during the COVID-19 crash, there were clear signals of market distress. Volatility smirk became significantly more steep, i.e. put options became relatively more expensive to call options during this period. Suppose that the values from this period are compared, for example, to the 75th percentile SKEW value of 4.7%, it becomes apparent that the volatility smirk was abnormally pronounced in the months preceding and during the COVID-19 crash. Although the volatility smirk of index options indicated some crashes, it did not predict them with certainty. Additionally, there appear to be some false signals that can be read from the figure. Therefore, additional variables must be considered to complement this promising detector.

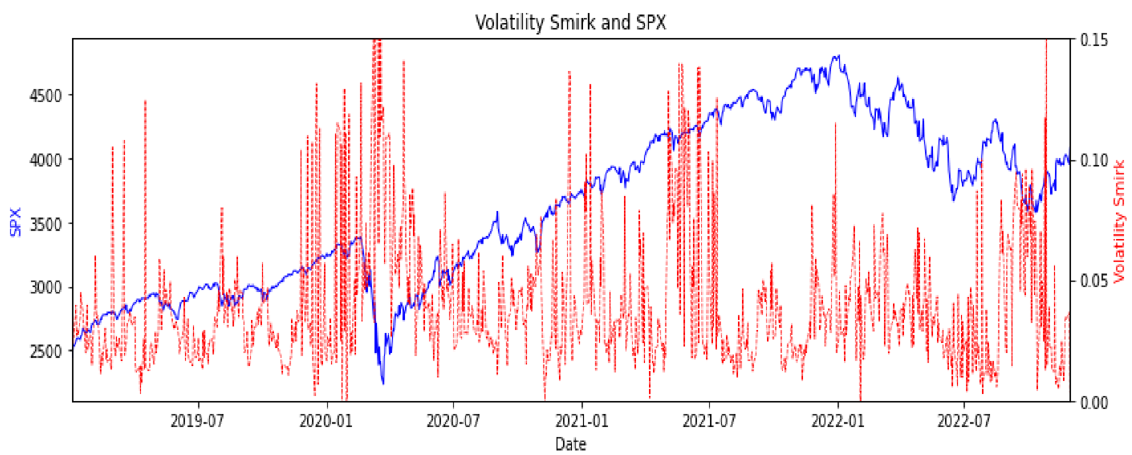


Figure 3. S&P 500 time series and the volatility smirk

4.2.2 Open interest difference

Open interest measures the total number of open option contracts in the CBOE market. The number value of open interest represents outstanding contracts that have not yet been settled, i.e., contracts that have not been exercised, closed, or expired. This metric provides insight into both the liquidity and trading activity in S&P 500 index options and the prevailing market sentiment surrounding it.

Call-put open interest difference denoted as $OIX_{i,t}$, is calculated as the difference between ATM S&P 500 index call option open interest, $OI_{i,t}^{ATMC}$, and ATM put option open interest, $OI_{i,t}^{ATMP}$, at matching days till expiration between 5 and 60. Similarly as before, the ATM option is defined as options with moneyness between 0.95 and 1.05.

$$OIX_{i,t} = OI_{i,t}^{ATMC} - OI_{i,t}^{ATMP}\ , \quad t = 5, \dots, 60, \tag{4}$$

Open interest difference indicates the overall market sentiment regarding the underlying. A negative call-put open interest difference suggests a negative sentiment towards the underlying as there is higher interest in put options than call options, as can be observed, for example, at the beginning of the pandemic in Figure 4. OIX has a mean of 252.21, ranging from -5,704.06 to 5,074.85, with a slightly negative skewness of -0.37. The mean of OIX is positive, indicating that during the four-year period, market sentiment has been more optimistic than pessimistic on average.

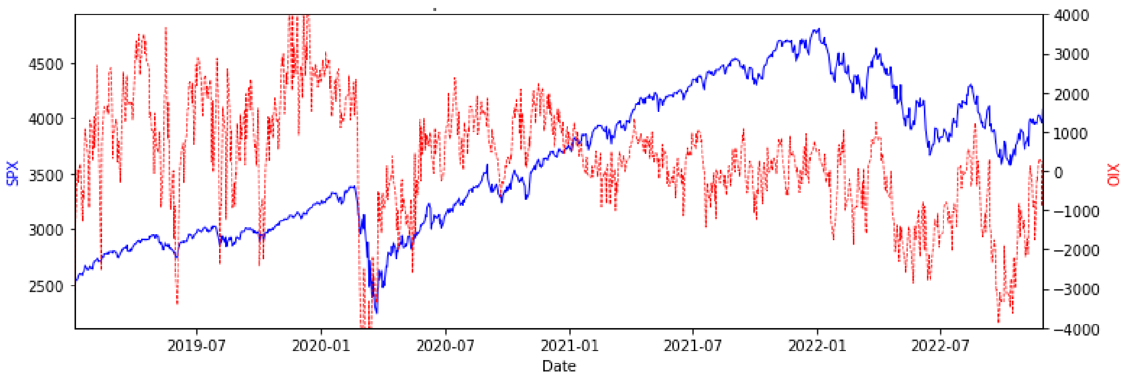


Figure 4. S&P 500 time series and the open interest difference (OIX)

The rationale behind using OIX as a potential detector of market crashes is straightforward. As documented before, most of the put option price derives from jump risk premia. Thus, when investors anticipate a market decline, they may choose to buy put options to protect their portfolios or make speculative profits. This behavior can lead to a surge in open interest in put options relative to call options, resulting in a negative OIX value. Intuitively, one might assume that a negative correlation between the OIX value and volatility smirk is too strong, potentially leading to multicollinearity. However, the later correlation matrix reveals that the correlation between the two is only -0.24. This suggests that multicollinearity is not a concern, and the two variables can complement each other, providing additional information that one variable alone may not capture.

4.2.3 Bid-offer spread

The final S&P 500-related option metric used in this study is the put option bid-offer spread, $SPREAD_{i,t}$, calculated as the difference between the offer price of the at-the-money put option, $OFFER_{i,t}^{ATMP}$, and the bid price of the same put option, $BID_{i,t}^{ATMP}$, divided by the offer price. The offer is the price at which a market participant can buy an option, while the bid is the price that one can sell an option.

$$SPREAD_{i,t} = \frac{(OFFER_{i,t}^{ATMP} - BID_{i,t}^{ATMP})}{OFFER_{i,t}^{ATMP}}\ ,\ t = 5, \dots, 60, \tag{5}$$

This variable serves as an excellent indicator of the instrument's liquidity. The rationale behind using it as a detector for market crashes is that market-makers might react by widening the spread before a volatility event. The spread is a variable that may detect subtle changes in liquidity before the two metrics discussed earlier come into play. Since market-makers are responsible for significant market volume, their actions would immediately affect the bid-offer spread.

Additionally, it is worth considering that leading market-makers, such as Optiver and Citadel, already possess and further heavily invest in state-of-the-art volatility models. As such, they may be the first parties to detect a potential market decline. As non-directional traders, they can process and react to this information in at least two ways: first, by withdrawing from the market, and second, by widening their spreads to reflect the new information. The latter is more probable as market makers are often committed to providing liquidity to the markets no matter the circumstances.

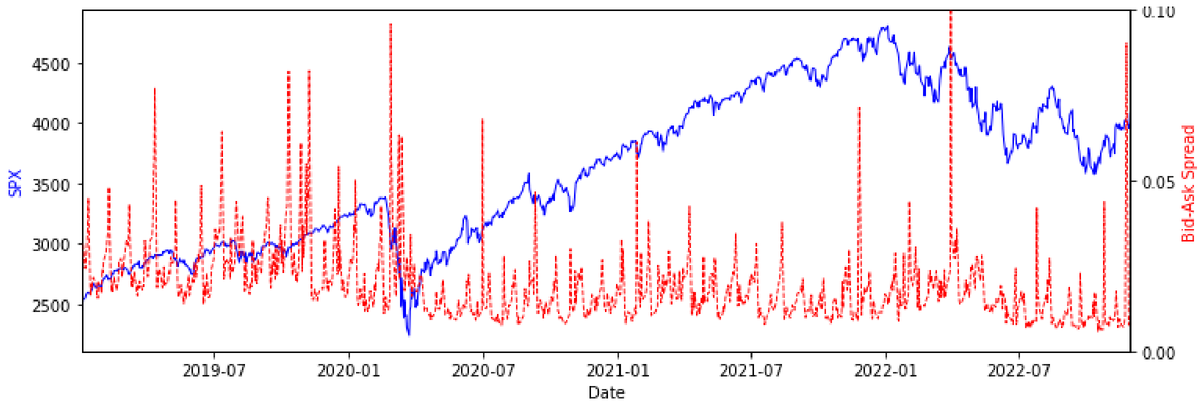


Figure 5. S&P 500 time series and the ATM put option bid-ask spread

Figure 5 shows that there are some trend regimes detectable, which also, together with the high volatility of the variable, leads to the Augmented Dickey-Fuller test being rejected only at a 10% level, meaning that there is not enough evidence of stationarity of the variable. As such, for the logistic regression, the variable is first differenced to remove seasonality. Calculation (5) below demonstrates the process of first differencing: $\Delta X_{i,t}$, represents the difference between the current value, P_t, and the previous value, P_{t-1}, of the variable, *i*, in the time series.

$$\Delta X_{i,t} = P_{i,t} - P_{i,t-1}, \tag{6}$$

In addition to the bid-offer spread, this operation is applied to the later specified variables. However, this Chapter introduces all variables for inspection purposes before any processing is made.

Notably, the put spread transferred from a higher spread regime to a slightly lower spread regime following the pandemic crash of 2020 and stayed in this regime for the rest of the period, excluding the extreme volatility spikes. Visual inspection of this variable does not yield such promising results, and the spikes in the spread seem somewhat random. However, the spikes are often very pronounced, and some patterns might not be evident in visual inspection.

4.2.4 Ted spread

TED spread is a financial metric that reflects the credit risk in the economy by comparing the difference between the interest rates level that banks lend to each other and the rate that government is capable of borrowing money. It is calculated as the difference between the USD-based three-month LIBOR, $r_{a,t}$, and the U.S. three-month T-Bill, $r_{i,t}$, rate.

$$TED\_Spread = r_{a,t} - r_{i,t}, \quad (7)$$

As LIBOR has been discontinued, the main replacement in the U.S. is The Secured Overnight Financing Rate (SOFR). However, this study still uses LIBOR as it was available in Refinitiv on a daily basis, calculated and published by the Intercontinental Exchange Benchmark Administration (IBA) for the time period required for this study.

The TED spread could be used as an indicator for a market crash because it can reflect credit risk and overall confidence in the financial system. As the spread widens, it signals increased perceived credit risk and decreased confidence among the market participants. In extreme cases, the increase in TED spread could lead to a contagion effect, where fear increases, leading to panic selling and declining equity indices.

The TED-Spread is usually narrow, on average only 0.26%, or 26 basis points, during the study period. This variable also has some volatility, with a value range of 1.44% and a standard deviation of 0.2%. Despite the volatility, the variable rejects the ADF test but is later in Chapter 4.4 found non-linear with the log odds. Thus the variable is first-differenced for the final regression model.

Figure 6 reveals that perceived credit risk rose sharply during the pandemic. After the worst fear period of the pandemic, the credit risk level flattened quickly. Consequently, TED spread variations were almost non-existent until the beginning of 2022. In the first quarter of 2022, even negative values were received as T-Bill rates rose due to contractionary monetary policy. After this, the spread remained volatile for the rest of the study period.

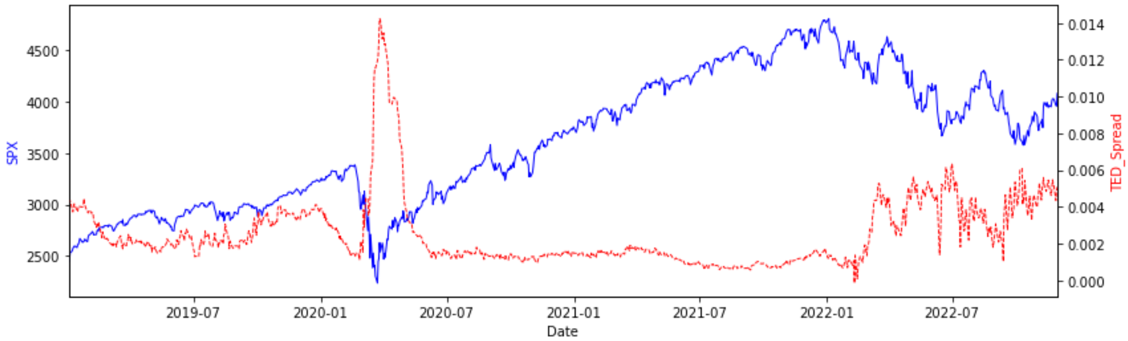


Figure 6. S&P 500 time series and the TED spread

It is known that the volatility of the rates in 2022 and 2023 was one of the main factors that led to the significant financial distress of multiple large banks. The prolonged distress was then followed by collapsing or restructuring of operations of the financial institutions. Banks such as Credit Suisse, Silicon Valley Bank, and First Republic no longer exist in the 2nd quarter of 2023.

4.2.5 Bond-stock earnings yield differential

The bond-stock earnings yield (BSEYD) is calculated as the earnings, $E_{i,t}$, divided by price, $P_{i,t}$, of S&P 500, *i*, minus a 3-month Treasury bill (USD), $r_{i,t}$. Refinitiv provided a daily price-to-earnings ratio of the S&P 500 and the middle rate of the 3-month Treasury bill, representing the rate at which the security is traded in the secondary market. As the variable presents seasonality, it is first-differenced before logistic regression, following calculation (5) to make it a suitable variable for the model.

The variable adheres to the methodology established by Lleo and Ziemba (2017). However, it reverses their approach as they calculated the variable by subtracting the earnings yield of the S&P 500 from the risk-free rate. This adjustment is made for better interpretability and intuition of the variable. While this adjustment does not impact the outcome of the variable, it does inverse the interpretation of the coefficient relative to their research.

$$BSEYD_{i,t} = \frac{E_{i,t}}{P_{i,t}} - r_{i,t}, \tag{8}$$

The T-Bill rate is an essential indicator of the creditworthiness of the U.S. government. As such, it is liquid and closely monitored by investors and policymakers as well. The S&P 500 price-to-earnings, which is inversed to find the earnings yield for the model, is a valuation ratio used to evaluate the index's current price compared to the earnings of the index, representing the sum of all 500 companies' earnings per share (EPS).

BSEYD measures how attractive returns in equities are compared to bond returns. If the variable has low values, it signals that there might be downward pressure in equity prices. Vice versa, if the variable has high values, it indicates that equities may be increasingly attractive, decreasing the likelihood of a market correction. In other words, a low BSEYD implies that investors may find bonds more attractive than stocks, leading to a potential decline in stock prices. The downward trend in 2022 can be detected from Figure 7 can be explained mainly by the aggressive interest rate policy that FED adopted to combat record-high inflation.

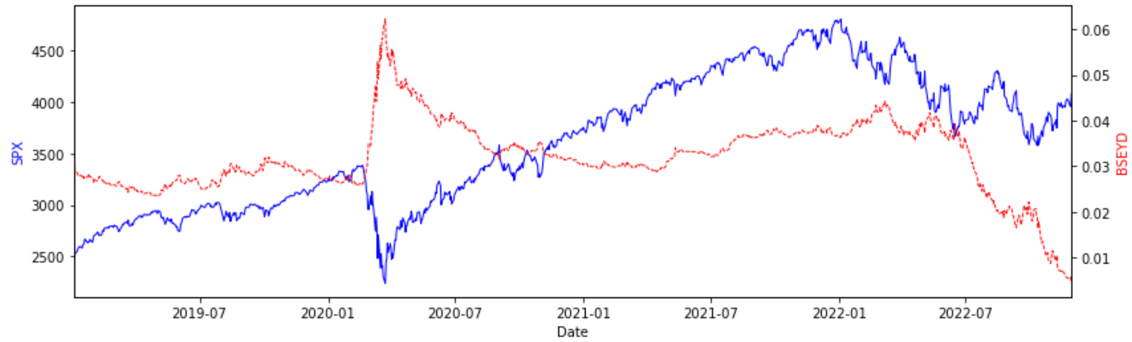


Figure 7. S&P 500 time series and the bond-stock earnings yield differential

For example, on October 31, 2022, S&P 500 price-to-earnings was 20.3, giving an earnings yield of 4.93% for the index. At the same time, U.S. 3-month T-Bill yielded 4.06% on an annualized basis. As a result, BSEYD received a low value of 0.87%, making Treasury instruments relatively attractive. For the total time span of the study, BSEYD has a mean of 3.20% and ranges from 0.53% to 6.23%. The standard deviation is 0.69%.

4.2.6 Term spread

The yield curve, also known as the term spread, is widely regarded as a strong leading economic indicator with a proven history of accurately predicting recessions. Generally, when the yield curve is negative or inverted, it suggests that short-term lending is more expensive than long-term lending, which indicates a looming recession. Inversion of the yield curve is somewhat atypical, as investors should expect higher returns for taking on the additional risk of holding the bond for a longer period. However, the inversion has occurred multiple times already in this millennium.

Prior research presents no clear consensus on whether the variable consistently provides insights into short-term future equity returns, irrespective of the curve's shape. As shown in Figure 8, the yield curve exhibited a variety of forms during the study period, ranging from steep and flat configurations to two instances of inversion. To address the potential seasonality of the variable, the yield curve is first differenced before incorporating it into the regression model.

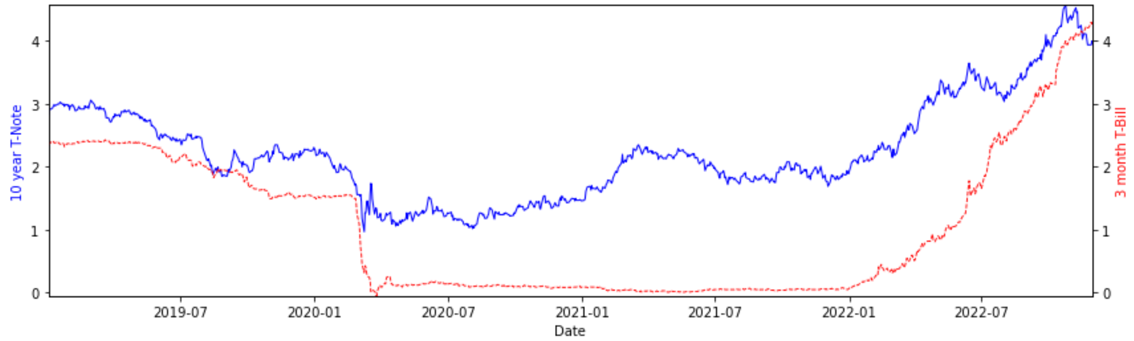


Figure 8. T-Bill and T-Note yields

The yield curve is calculated as the difference between the 10-year Treasury note yield, $r_{n,t}$, and the 3-month U.S. Treasury bill rate, $r_{i,t}$:

$$Yield\ curve = r_{n,t} - r_{i,t}, \tag{9}$$

Analyzing the yield curve and its changes over time can give insight into the market's expectations of future interest rates and economic outlook. During the study period, the yield curve ranged between -0.36% and 2.54%, with a standard deviation of 0.79%. Figure 8 displays constituents of the term spread, 3-month T-Bill yields with red and 10-year T-note yields with blue. Displaying the instruments separately may provide some additional insight into the behavior of the term spread.

Figure 9. S&P 500 time series and the yield curve

Following the Global Financial Crisis in 2007-2008, E.U. and USA exercised relatively expansionary monetary policy by keeping the interest rates historically low. After the pandemic, governments and central banks reacted even more aggressively to revive the shutdown economy by further lowering interest rates. After a period of quantitative easing, inflation skyrocketed, hence the increase in the yields at the beginning of 2022. The yield curve has inverted twice during the study period when examining the yields of long- and short-term lending in Figure 9 The first inversion took place in the last quarter of 2019, and the second occurrence was at the end of the study period in 2022. Whether the yield curve gives information about short-term returns and market declines cannot be determined after visual inspection.

4.2.7 Inflation

This paper incorporates the S&P GSCI Commodity Total Return Index as a variable to account for the relationship between equity and commodity returns. This index, widely recognized as the leading measure of general commodity price movements and inflation in the global economy, is continuously updated and comprises 24 commodities (S&P Global, 2023). These commodities are weighted according to their global production quantities, with the largest weight assigned to the commodities produced in the greatest volumes, with an annual recalibration. As a total return index, it reflects not only the price levels of futures contract prices but also the collateral yield. The index uses a monthly rolling methodology, purchasing the most liquid near-term futures contracts following the aforementioned weighting scheme.

For a financial time series like S&P GSCI, simply differencing or taking a natural logarithm of the variable is not enough to achieve stationarity. In such cases, both differencing and log transformation is necessary to remove exponential trends and other non-stationary components in the data.

Log difference, $\Delta Log_{i,t}$, is calculated by taking the difference between the natural logarithm of the variable at time t and time t-1. After this process, the series becomes stationary and thus is suitable for further modeling.

$$\Delta Log_{i,t} = \log(P_{i,t}) - \log(P_{i,t-1}), \quad (10)$$

4.2.8 Economic activity

The final variable studied in this paper is The Baltic Exchange Dry Index (BDI), a leading indicator of global commodity demand and economic growth. Papailias et al. (2017) summarize that the BDI is calculated as a daily weighted average freight price to ship raw materials across 20 different routes. Consequently, it incorporates information on future economic activity regarding the supply and demand for the transported materials.

To reflect on past events, BDI seems to capture the state of the economic cycle. For example, BDI showed record low values already preceding COVID, and at the end of 2021, it signaled that the economy would be close to the peak of the cycle by presenting extremely high values with a subsequent decrease that followed.

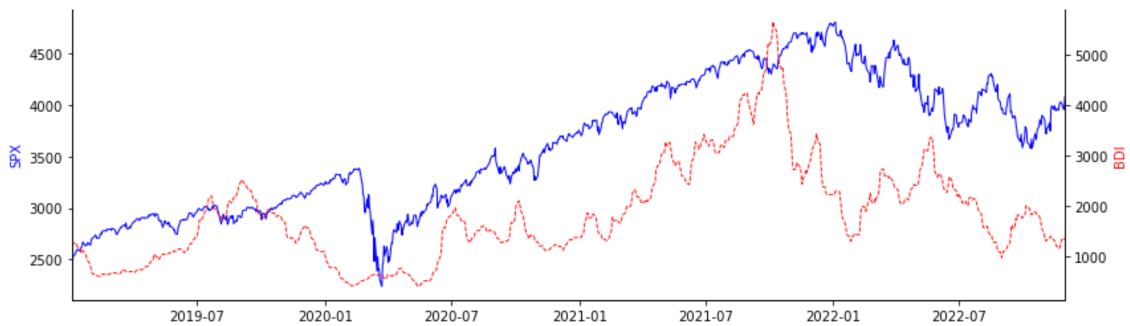


Figure 10. S&P 500 time series and the baltic dry index

Based on the findings and framework provided by Bakshi et al. (2010), BDI is calculated as log changes in the index over the preceding month. In the calculation, 21 trading days represent a lag of 1 month. As per visual inspection from Figure 10, the variable seems to display some degree of indication of the future direction of equity returns.

$$BDI_{growth} = \log\left(BDI_t - BDI_{t-21}\right), \tag{11}$$

4.3 Descriptive statistics

This section presents the descriptive statistics of the processed variables. An additional step taken to refine the data involves lagging each independent variable by one to remove any potential look-ahead bias. Consequently, the model is left with a total of 964 observations.

Table 4 below displays the same variables as Table 2 and in the same sequence discussed above. However, the variables are shown with their final, processed, and stationarized values in this case. Variables that have been differenced are marked with Δ, and those that have been log differenced are marked with Δlog. As a result, many of the variables exhibit low mean and standard deviation values.

Compared to unprocessed variables presented in Chapter 4.2, Table 4 introduces variables after processing. In addition, it introduces a new variable, *Crash*, which is the dependent variable in this study, representing a binary classification of market conditions. A value of 0 indicates that a crash is occurring based on the criteria introduced in Chapter 4. Otherwise, when a market crash is not occurring, the variable is assigned a value of 1. The mean of 0.711 for the Crash variable signifies that, within the 964 observations collected over the four years, 71.1% (or 686 observations) are assigned a value of 1, indicating that the market is going up. The remaining 28.9% (or 278 observations) are assigned a value of 0, representing crash or market decline occurrences.

Table 4. Descriptive statistics of the processed variables

	Mean	SD	Range	Min	Max	Kurt	Skew
Crash	0.71	0.45	1.00	0.00	1.00	- 1.03	- 0.98
Skew (%)	4.00	2.88	22.58	-1.91	20.67	4.96	1.96
OIX	253.73	1 470.86	10 778.91	-5 704.06	5 074.85	0.71	- 0.37
Δ Spread (%)	0.00	1.10	16.52	-8.27	8.25	20.67	- 0.68
TED_Spread (%)	0.26	0.20	1.44	-0.02	1.42	9.82	2.66
Δ BSEYD (%)	- 0.00	0.08	1.26	-0.53	0.73	14.27	0.96
Δ Yield curve (%)	- 0.00	0.06	0.69	-0.32	0.37	5.13	- 0.03
Δ log_S&P GSCI	0.00	0.02	0.20	-0.13	0.08	9.15	- 1.22
Δ21 log_BDI	0.00	0.33	2.05	-0.78	1.27	0.99	0.33

Examining the table, *OIX* and *BDI* have relatively large values, which could warrant scaling for improved comparability of the coefficients. However, scaling does not alter the underlying relationships between the variables or the overall model's validity. Therefore, in this case, the two variables are not scaled, and instead, coefficients are carefully interpreted while considering their magnitudes.

4.4 Logistic regression

This chapter includes an introduction to the methodology and mathematical framework surrounding it. As specified before, this study conducts a logistic regression on the dependent variable *Crash* and the independent variables introduced in Chapter 4.1. Logistic regression employs a binary logic, meaning that a market crash either occurs or does not, making the model non-linear but straightforward to interpret.

In logistic regression, the logistic function (12) is used. The outputs of this function will be discussed in detail in Chapter 5. The logistic function, *p(X),* gives outputs between 0 and 1 for all values of X. For low values of X, the function predicts that the probability of a market crash is high. Vice versa, when the function displays values closer to 1, it indicates that selected variables reflect a low possibility of a market crash. The logistic function always produces an S-shaped curve (James et al., 2021).

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_k X_k}}, \tag{12}$$

After manipulating calculation (12) above, *odds* can be found. Odds can take any value between 0 and ∞. Values close to zero indicate a high probability of a market crash, whereas values closing ∞ indicate that the probability of a market crash is low, i.e., the probability of the market appreciating is high. This may seem counter-intuitive, but this stems from the definition of the dependent variable, *Crash*, which was defined so that a value of 0 indicates a crash, and a value of 1 indicates no crash.

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X_1+\ldots+\beta_k X_k}, \tag{13}$$

Where, $p(X),$ is the probability of the event (in this case, a market crash), $\beta_0 + \beta_1 + \ldots + \beta_k$, are the coefficients of the model (to be estimated from the data) $X_1 + \ldots + X_k$, are the predictor variables introduced in Chapter 4.2.

Applying the natural logarithm to the equation (13) will result in log-odds or logit, represented by the left-hand side of the equation (14) below. The equation shows that increasing X_k by one unit changes the log odds by β_k. When β_k is positive, increasing X is associated with an increase in p(X). Conversely, when β_k is negative, increasing X is associated with a decrease in p(X).

$$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k, \tag{14}$$

Equation (14) demonstrates that logistic regression assumes a linear relationship between independent variables and the log odds ratio (James et al., 2021). This relationship is visually tested in the next chapter to ensure this assumption holds with the used variables.

4.5 Validation of the model assumptions

To ensure the credibility and reliability of the model, its underlying assumptions must be validated. Although logistic regression does not inherently require the predictors to be stationary, the specific nature of this requires it. Given that the dependent variable is a binary indicator of a market crash derived from the S&P 500 - a time series - all variables will be subjected to the Augmented Dickey-Fuller (ADF) test to check for stationarity. This precaution aims to avoid potential spurious results that could arise from coincidental time trends in the independent variables, which may correspond with trends in the dependent variable.

$$H_{o,ADF}: A\ unit\ root\ is\ present\ in\ the\ time\ series.$$
$$H_{a,ADF}: A\ unit\ root\ is\ not\ present\ in\ the\ time\ series$$

ADF tests the null hypothesis, $H_{o,ADF}$, that a unit root is present in the time series of the variable. The null hypothesis, $H_{o,ADF}$, is rejected if the test statistic is more negative than the critical value. The alternative hypothesis, $H_{a,ADF}$, is that the time series of the variable is stationary. A unit root suggests that the time series is non-stationary, implying that the mean and variance of the series are not constant over time. If such a case arises, it may lead to spurious regression results. All testing was conducted via statistical software.

Table 5. Augmented Dickey-Fuller test

	Test Statistics	Critical Values at 1% significance
Skew	- 6.47	- 3.44
OIX	- 4.89	- 3.44
ΔSpread	- 14.10	- 3.44
ΔTED_Spread	- 6.46	- 3.44
ΔBSEYD	- 7.06	- 3.44
ΔYield_curve	- 7.06	- 3.44
Δlog_S&PGSCI	- 30.21	- 3.44
Δlog_BDI	- 4.14	- 3.44

The ADF test results in Table 5 show that the presence of a unit root is rejected for all variables at a 1% significance level. Given that the number of observations is the same for each variable, the critical value remains consistent across all variables. This allows the conclusion that the variables' time series are indeed stationary.

Following the stationary check, the logistic regression model's assumptions are examined as outlined by Stoltzfus (2011). To ensure the model's reliability, the subsequent assumptions must be checked: (1) Independence of observations, (2) Linearity of independent variable and log-odds, (3) Absence of multicollinearity, and (4) Absence of strong influential outliers. Stoltzfus also notes that there should ideally be between 10 to 20 observations per covariate in the model. The criterion is satisfied by the model, there is more than (8*20) 160 observations in the study.

The first assumption means that the observations should not come from repeated measurements or matched data. This assumption is considered to be met by the design of this study; the data was collected from the financial markets, where each data point is independent of the others. Therefore, there is no inherent relationship or dependence between observations in the dataset.

Even though the logistic regression model does not require a linear relationship between dependent and independent variables, it does assume linearity between log odds and independent variables, as indicated in the equation (14). Thus, before logistic regression can be performed, the linearity to log odds of these variables must be analyzed.

This is a critical assumption for logistic regression, and with TED_Spread, it is initially not fulfilled. Figure 11 shows the original relationship with TED_Spread and log odds, while Figure 12 represents the final relationships after modification. As Figure 11 reveals, initially TED_Spread did not fulfill the linearity condition with log odds; the relationship is better described as parabolic or U-shaped

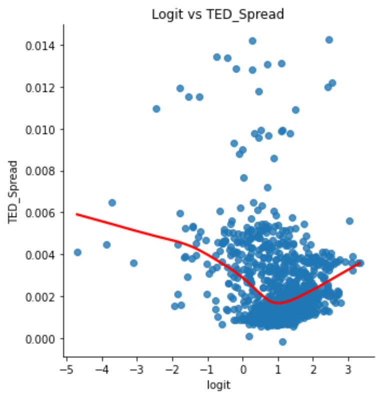


Figure 11. The TED spread and violation of log-odds linearity

As a result, multiple methods were implemented to find a remedy to this issue. The application of natural logarithm failed to linearize the relationship, whereas adding a polynomial term led to convergence issues in the model. The use of a log-difference transformation, while resulting in linearity, diminished the predictive power of the TED_Spread variable substantially.

Ultimately, simply first differencing the variable was the most successful approach. It achieved linearity with log odds but also preserved some predictive power of the variable, despite a noticeable reduction. However, the significance of the variable moved from the original level of 1% to 5%, which also diminished the predictability power of the whole model. The final relationship together with all other variables, post-transformation, is presented in the following figure.

Figure 12 represents the log odds relationship to each variable. The smoothed scatter plots in the Figure show that, after the modification, all variables exhibit a satisfactory linear relationship with the 'crash' outcome in the logit scale. Visual inspection is the primary tool for this assumption check.

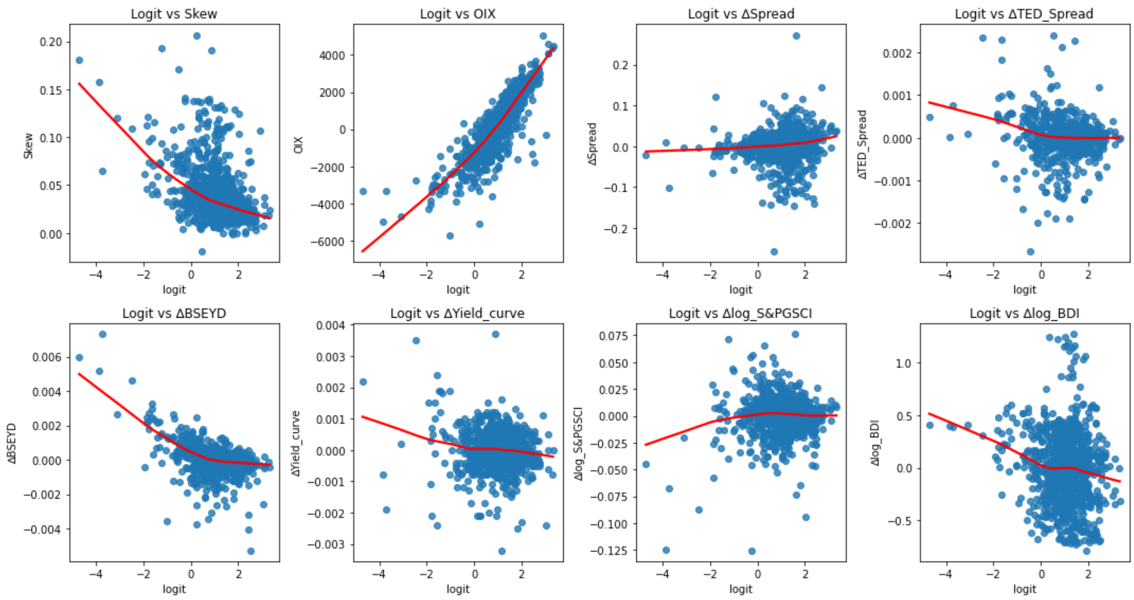


Figure 12. Linearity of the variables with log-odds

The third assumption of the logistic regression is the absence of multicollinearity. Correlation matrix in Figure 13 is used to detect possible multicollinearity among the independent variables as well as to analyze the relationships among them. It shows values between -1 and 1, where -1 indicates a perfect negative relationship and 1 perfect positive one, respectively. There should not be any correlations that are close to either extreme, as such values would signify multicollinearity that could inflate the variance of the estimated coefficients in the regression model and thus make it unreliable.

Overall, the correlation matrix shows that most variables have a weak relationship with each other, which is a desirable finding. Some of the strongest correlations are between OIX and Volatility smirk, Yield curve, and TED spread and with BSEYD and S&P commodity index. The strongest (positive) relationship is between BSEYD and TED spread.

The correlation matrix is used for further examination of the variables. For example, it is no surprise that OIX and Skew exhibit a slightly above-average negative correlation as they both measure market sentiment from different angles. A closer examination of this relationship suggests that if Skew increases in value, i.e., put options prices increase relative to call option prices, then Open Interest in call options relative to put options typically decreases, and vice versa. An increase in Skew and a simultaneous decrease in OIX would both signify a shift towards bearish market sentiment.

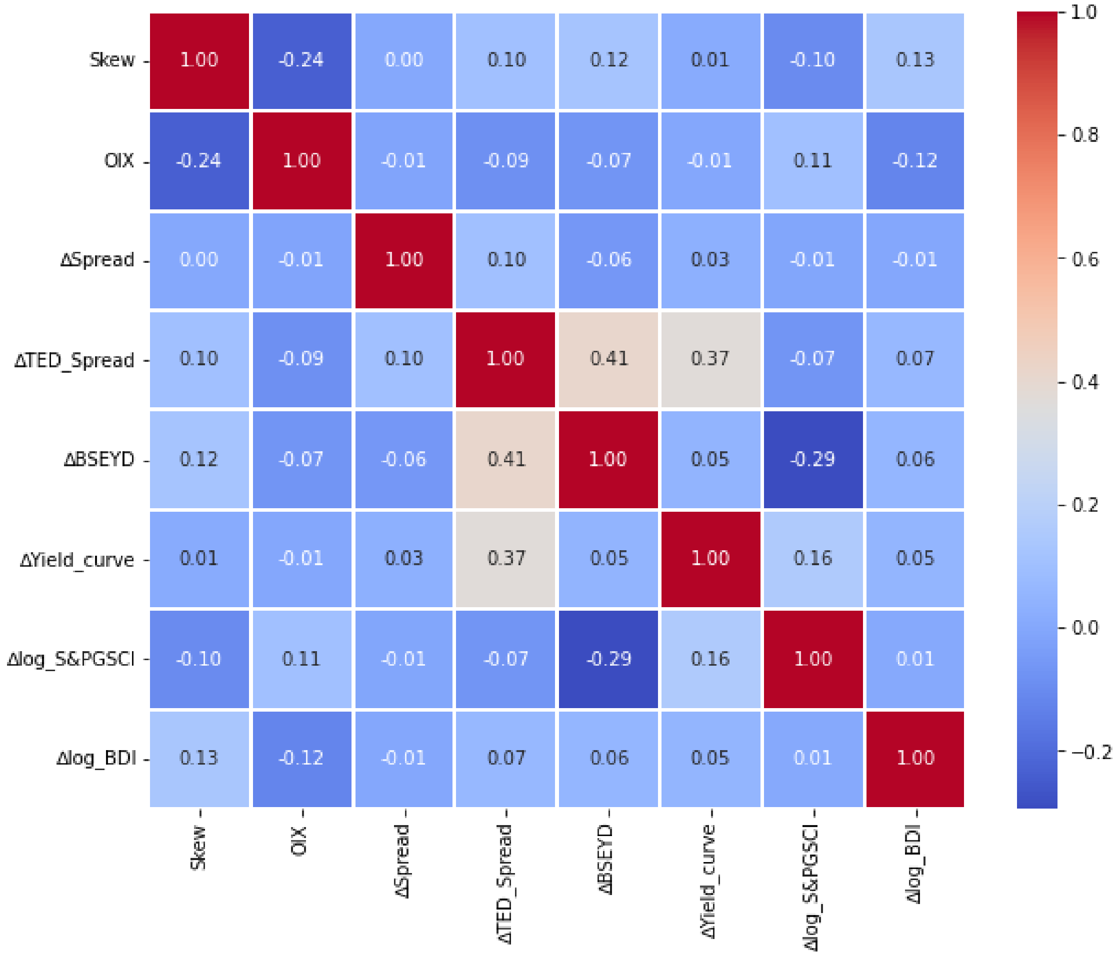


Figure 13. Correlation matrix

The correlation matrix helps verify whether these relationships align with the expectations and understanding of the underlying economic or financial relationships. Any unexpected correlations may signal the need for further investigation, as they could, for instance, reflect data issues, or provide insights about the market dynamics.

The Variance Inflation Factor (VIF) is a more scientific measure of multicollinearity, calculated as the ratio of the variance of the overall model to the variance of a model that includes only a single independent variable. A VIF value of one is the minimum value and represents no collinearity, while a value greater than five can be considered problematic. To conclude, analyzing the correlation matrix, together with Variance Inflation Factors, it is certain that there is no multicollinearity.

Table 6. Variance inflation factors

	VIF
Skew	2.13
OIX	1.02
ΔSpread	1.01
ΔTED_Spread	1.45
ΔBSEYD	1.14
ΔYield_curve	1.04
Δlog_S&PGSCI	1.15
ΔBDI	1.01

The final assumption to validate is the absence of strong outliers. Cook's distance (see, Cook, 2000) detects influential observations, while standardized residuals identify whether the observation is an outlier. Cook's distance measures how much the predicted values for the response variable would change if a particular observation were excluded from a linear regression model. The calculation is rather complex by hand, but it is done using statistical software. Glen (2023) summarizes that observation with a Cook's distance value above 0.5 can be considered influential, and warrant further investigation, especially if it is an outlier as well.

Standardized residuals represent the difference between the observed and predicted values divided by the standard error. A conservative rule of thumb is that the absolute values of over 2 represent an outlier. Table 7 presents 3 observations with the largest Cook's distance and their standardized residuals. Following the thresholds introduced above, it can be concluded that none of these observations is an outlier, nor are they influential. Therefore, there will be no influential outliers in the regression.

Table 7. Cook's distance and standardized residuals

Number of the Obsevation	Standardized Residuals	Cook's Distance
710	- 1.89	0.039
281	- 1.59	0.032
156	- 1.99	0.018

5 Empirical results

This Chapter unveils the empirical findings derived from the data and methodology previously outlined. Initially, the full sample logistic regression model is analyzed, and the results with all predictors are presented. Following this, the variables lacking predictive power are eliminated, and the remaining model is analyzed. This enables comparison and potential insights into the dynamics of the markets.

Subsequently, the full-sample results of the regression models are introduced and discussed. This discussion includes logistic regression models using the most robust variables while comparing thresholds of 3%,4% and 5% for the dependent variable. Additionally, the section demonstrates the respective models' model fit, accuracy, and capability using tools such as confusion matrices and area under the ROC curve.

The final discussion centers on the out-of-sample performance of the three models to enhance the robustness of the executed tests. The end goal is to assess and demonstrate the model's generalizability by providing a comprehensive understanding of its performance with data it has not seen before.

5.1 The logistic regression model, analysis of the variables

This chapter delves into the empirical results from two distinct models. Both models use the same dependent variable, 'Crash4Pct', which sets a 4% threshold for a market crash. It is designated as '0' if a crash occurs and '1' otherwise. The differentiation between the two models lies in the selective removal of the two variables that were found to be insignificant or did not produce anticipated results.

The independent variables used in the regression model were introduced in Chapter 4: volatility smirk (Skew), open interest difference (OIX), put option bid-offer spread (Spread), TED spread (TED_Spread), bond-stock earnings yield differential (BSEYD), yield curve (Yield_Curve), S&P GSCI commodity index (SPGSCI) and finally Baltic Dry Index (BDI). Table 8 presents the regression results in full. All interpretations are derived from model (2) unless otherwise noted.

Next, effect size and statistical significance of the variables are discussed. The effect size measures the magnitude of the relationship between the predictor, independent variable, and the outcome, dependent variable, and is expressed as a log-odds ratio. A larger log-odds ratio suggests that our predictor significantly influences the outcome. In contrast, a smaller ratio means the predictor has less impact: changes in the predictor don't substantially shift the outcome. However, the independent variable's scale and variation must also be considered before making any conclusions.

First, variables 'BDI' and 'Spread' are analyzed from the model (1), as they are subsequently removed from the model (2) due to their lack of predictive power. Following that, we explore the remaining variables through the lens of Model (2), distinguished by its relatively lower Akaike Information Criteria (AIC). A lower AIC value represents a better fit of the model to the data, as it balances model complexity (number of variables) against performance, thus implying an optimal trade-off. In other words, the removal of variables led to a lower AIC value for model (2), implying that it is superior to model (1).

Bid-offer spread is removed from model 2 due to low significance and inconsistent results. The variable's positive coefficient suggests that the odds of a market crash decrease when the bid-offer spread widens, indicative of decreasing liquidity in the market. This behavior contradicts the expectations set when selecting the variable. However, with the p-value of the variable being as high as 0.49, any interpretation made should be taken with caution. In conclusion, the statistical evidence suggests that the variable does not meaningfully contribute to the crash likelihood.

Similarly, BDI, an indicator of economic activity, exhibits a very high p-value of 0.90. Its positive coefficient suggests that when economic activity rises, 30 days later, the odds of a crash decrease. Although it captures the correct dynamics, the statistical significance is nonexistent, necessitating its exclusion from Model 2. So, while Bakshi et al. (2010) find a highly positive and statistically significant coefficient with BDI and stock market returns, this study did not find similar results.

Table 8. Logistic regression results, analysis of the variables

Results		
	Dependent variable:	
	Crash4Pct	
	(1)	(2)
Skew	-6.909***	-6.878***
	(2.667)	(2.654)
OIX	0.0005***	0.0005***
	(0.0001)	(0.0001)
Spread	1.554	
	(2.262)	
TED_Spread	452.468**	474.283**
	(222.868)	(221.001)
BSEYD	-552.878***	-560.915***
	(115.694)	(115.260)
Yield_Curve	-189.789	-190.532
	(143.699)	(143.636)
SPGSCI	-6.154	-6.187
	(4.790)	(4.782)
BDI	0.031	
	(0.234)	
Constant	1.149***	1.147***
	(0.136)	(0.135)
Observations	964	964
Log Likelihood	-514.000	-514.244
Akaike Inf. Crit.	1,046.000	1,042.489
Note:	*p<0.1; **p<0.05; ***p<0.01	

The Volatility smirk is significant at the 1% level in both models, exhibiting a negative coefficient of -6.88 with a standard error of 2.65. The coefficient suggests that a one-unit (one percent) increase in Skew corresponds to a change in the log odds of a market crash. More intuitively, a one-unit increase in the relative expense of OTM put options compared to ATM call options amplifies the odds of a market crash by 0.1%, calculated as exp(-6.88). This interpretation assumes all other variables remain constant. Translating the odds to probability, a 0.1% increase in the odds equates to a 0.1% increase in the crash probability, as per calculation (15). Given that Skew ranges from -1.91% to 20.67%, Skew alone can substantially impact the odds of a market crash.

$$\Delta Probability = \frac{odds}{1+odds}, \tag{15}$$

Open interest difference is another variable statistically significant at a one percent level. OIX has a positive coefficient of 0.0005, implying that every time OIX increases by one unit, the odds of a market crash decrease by exp(0.0005)= 0.05%. Conversely, when OIX decreases by one unit, the odds of a market crash increase by 0.05%. While this percentage may seem trivial, it is essential to remember that OIX has a wide value range extending from -5704 to 5075. As a result, it may be more suitable to look, for example, at a 100 units change in OIX from 1100 to 1000, which would have increased 5% in the odds of the crash, all else remaining constant. Skew and OIX are the only variables presented in absolute values, as the others are processed by either differencing or log differencing.

TED spread, a first-differenced variable, suggests that the coefficient describes the change in the log odds of the outcome for a one-unit increase in the change of the predictor, rather than a one-unit increase in the level of the predictor itself. Summary statistics of the first differenced TED_Spread are in Table 9 Given that the variable's standard deviation is 0.4%, a 1 unit increase is unlikely. As a result, with a positive coefficient of 421.41, 0.0001 unit increase in change of TED spread corresponds to (exp^(474.28*0.0001)) 4.9% decrease in the odds of a market crash.

Table 9. First differenced values of TED spread

	Mean	SD	Range	Minimum	Maximum	Kurt	Skew
ΔTED_Spread	0.0000	0.0004	0.0051	- 0.0027	0.0024	10.6968	- 0.0167

Even though the TED spread is statistically significant at the 5% level and included in Model 2, it exhibits a somewhat counterintuitive relationship with the dependent variable. As interpreted above, a widening TED spread actually decreases the odds of a market crash, contradicting the prevailing theory that a surge in the TED spread, indicative of increasing credit risk in the economy, would inherently heighten the likelihood of a market crash due to potential contagion effects spilling over into the equity markets.

One potential explanation for this unexpected behavior could be the timing or lag of the effect. It's plausible that increases in the TED spread may not have an immediate impact on the equity market, and the lag effect isn't captured well in the current model. Further, the anomalous behavior of the TED spread might be explained by its transformation into a first-differenced form. When a variable is first differenced, the focus shifts from the absolute level of the variable to its changes from one observation to the next. Thus, the TED spread in our model signifies changes in perceived credit risk, not the overall level experienced by the economy. In addition, as seen in Figure 6, TED spread had extremely low variation following pandemic that lasted until the beginning of 2022, thus yielding less information for the model.

The BSEYD variable is statistically significant across both models at the 1% level, exhibiting a negative coefficient of -560.915. This coefficient implies changes in the short-term risk-free rate or fluctuations in the earnings yield can impact the likelihood of a market crash.

Focusing first on the earnings yield through the price-to-earnings ratio (P/E). A rising P/E can be understood best through two scenarios: the index's price level increases while earnings remain constant, or earnings drop while the price level holds steady. Most likely, it is a combination of the two. However, when inspecting the variable, it should be noted that price movement fluctuations tend to be more volatile than the trajectory of earnings.

When the P/E ratio declines, the earnings yield rises, as it is the inverse of P/E. Assuming constant interest rates, a surge in the earnings yield would consequently lead to an increase in the first differenced BSEYD. To underscore this impact, we can consider a hypothetical increase in the BSEYD by 0.0001 units. In such a scenario, the odds of a market crash rise by approximately 5.45%. This may be derived not only from price momentum but also from earnings momentum. That is, rapid changes in earnings could also contribute to fluctuations in BSEYD and thus could be an additional factor influencing the probability of a market crash.

On the other hand, when risk-free interest rates rise, the yield on new bond issuances also increases, making them more attractive to investors. This could lead to a sell-off in equities as investors might want to increase the weighing of bonds in their portfolios. This reallocation could counterbalance the increasingly more attractive equity prices witnessed at the end of 2022.

The Yield Curve variable is not statistically significant in the regression, as evidenced by its p-value of 0.187, which exceeds the conventional significance thresholds. The coefficient of the Yield Curve is negative, suggesting that the odds of a market crash increase when the spread between long-term and short-term interest rates widens. This aligns with Figure 9 visualizations and the rationale for including the Yield Curve variable in the model.

Typically, an inverted yield curve is considered a reliable leading indicator of a recession. However, a recession does not typically occur during the inversion. Instead, it often happens once the spread begins to steepen again: short-term rates decrease faster than long-term rates or long-term rates increase faster than short-term rates. This is visualized in Figure 14, where the grey vertical areas represent economic recessions.

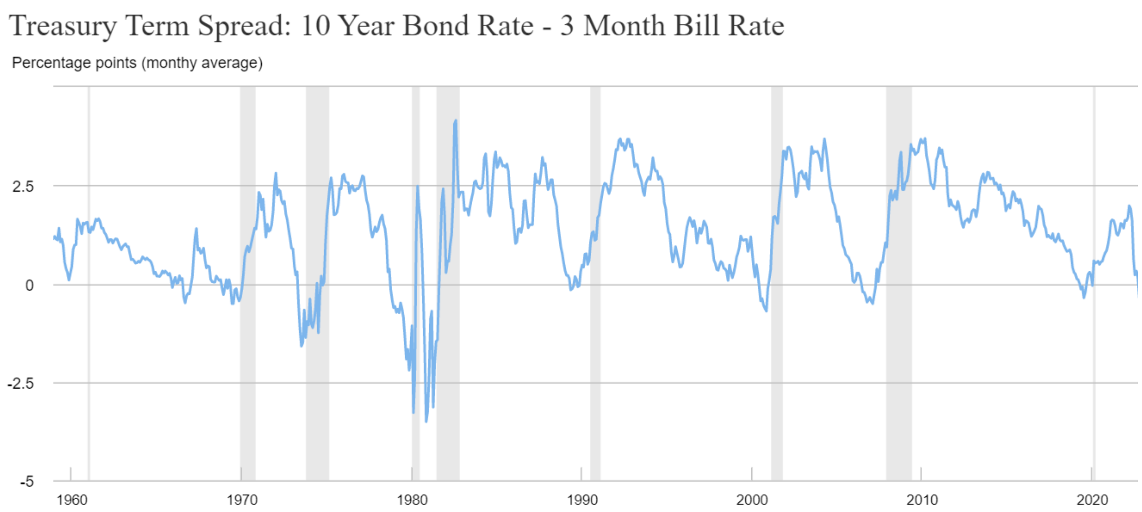


Figure 14. Historical relationship of the yield curve and U.S. recessions (Federal Reserve Bank of New York)

In this context, the negative coefficient for the Yield Curve variable could be capturing this phenomenon. As the yield curve steepens following an inversion, the odds of a market crash increase. However, a p-value of 0.187 suggests that while this variable may have some predictive power, it is not a strong determinant within the context of this model.

As the variable is first-differenced, it largely ignores the absolute level of the variable and emphasizes the change in it. Indeed, the absolute levels of variables such as BSEYD, TED spread, bid-offer spread, and the Yield Curve could hold relevance in maintaining predictive power. It is not just the fluctuations but also the absolute levels of these variables that could play a vital role in accurately capturing their dynamics with the market. Efforts were made to represent the variables in this manner; however, this approach led to a violation regression assumptions.

Commodity prices, represented by the S&P GSCI commodity index (SPGSCI), demonstrate an inverse correlation with the odds of a market crash, possessing a p-value of 0.19. In alignment with theoretical expectations, an increase in inflation mirrored through rising commodity prices, correlates negatively with equity returns. Despite its statistical insignificance, similarly to the yield curve, the SPGSCI variable is retained in the second model. The justification for this lies in two key areas: first, commodity prices reflect crucial economic dynamics; second, a substantial body of academic research advocates for the inclusion of this variable. Consequently, despite its limited statistical significance in the current model, the SPGSCI index retains a potential role in forecasting market corrections.

5.2 Performance analysis: full sample regression models

This section takes a closer examination of the overall performance of the refined regression models. Additionally, two new models are introduced, utilizing alternative thresholds of 3% and 5% for the dependent variable.

Initially, the statistical results derived from these models are assessed. Subsequently, a deeper exploration into their predictive performances is undertaken, employing tools such as confusion matrices and the Area Under the Receiver Operating Characteristic (ROC) curve. The models discussed herein incorporate only the variables that were rationalized as essential and informative in Chapter 5.1.

The overall fit and statistical significance of the models, as presented in Table 10, are assessed using Chi-squared statistics and their corresponding p-values. Regression (1), 'Crash3Pct', defines the dependent variable with a threshold of 3%, regression (2), 'Crash4Pct', presents previously observed results for the threshold of 4% and finally, regression (3), 'Crash5Pct', on the right-hand side employs a threshold of 5%.

Table 10. Comparative logistic regression results, for models (1), (2) and (3)

Results

	Dependent variable:		
	Crash3Pct	Crash4Pct	Crash5Pct
	(1)	(2)	(3)
Skew	-8.648***	-6.878***	-7.240***
	(2.573)	(2.654)	(2.774)
OIX	0.0004***	0.0005***	0.001***
	(0.0001)	(0.0001)	(0.0001)
TED_Spread	455.358**	474.283**	281.136
	(216.139)	(221.001)	(222.874)
BSEYD	-530.428***	-560.915***	-491.340***
	(112.480)	(115.260)	(115.798)
SPGSCI	-1.838	-6.187	-5.653
	(4.672)	(4.782)	(4.954)
Yield_Curve	-132.705	-190.532	-66.425
	(139.274)	(143.636)	(147.017)
Constant	1.045***	1.147***	1.488***
	(0.131)	(0.135)	(0.145)
3% Model Chi-sq (overall)	120	Overall p-value	1.7e-23
4% Model Chi-sq (overall)	130	Overall p-value	1.5e-25
5% Model Chi-sq (overall)	123	Overall p-value	3.6e-24
Observations	964	964	964
Log Likelihood	-544.714	-514.244	-463.432
Akaike Inf. Crit.	1,103.429	1,042.489	940.863
Note:			*p<0.1; **p<0.05; ***p<0.01

Model (2) has the highest Chi-square value of 130, indicating the most robust rejection of the null hypothesis, H_0 . However, the p-values associated with these Chi-squared statistics are practically zero, significant at the 1% level for all models. The low p-values provide strong evidence against the null hypothesis, H_0 , thus alternative hypothesis, H_a, is accepted

$$H_0: \textit{Predictors at time 't', have no effect on the likelihood of a crash at time 't} + 1'$$

$$H_a: \textit{At least one of the predictors at time 't', have an effect on the likelihood of}$$

$a\ crash\ at\ time\ 't+1'$

These results confirm the statistical significance of models' ability to predict the probability of a crash. It is important to note that the Chi-Square test does not directly measure the model's fit to the data, unlike some of the subsequent tests discussed. Instead, it only shows how confidently the null hypothesis, H_0, can be rejected. It seems that the higher threshold for defining a crash is used, signs of a crash become more robust and easier to detect. Intuitively, a higher threshold should lead to a more robust rejection of the null hypothesis, as the larger the market fluctuations could have increased preceding symptoms in the market together with less noise in the data.

Conversely, the log-likelihood measures how well the model replicates the observed data, serving as a key indicator of the model's fit. Therefore, even though a model may show a significant Chi-square value, implying at least one predictor has statistical significance, it might still exhibit a less-than-optimal log-likelihood value. In other words, despite statistically significant predictors, the model may not accurately or reliably predict the outcome variable due to potential issues like overfitting, underfitting, or an inability to account for all the complexities in the data. The closer the value approaches zero, the better the model fits the data. Thus, Model (3) has the best fit to data by having lowest log-likelihood value of -463.4.

The Akaike Information Criterion (AIC) evaluates the quality of the regression model by considering both the model's fit, or how much of the variation it explains, and the complexity of the model, represented by the number of independent variables it has. A lower AIC value denotes a superior model. According to this criterion, as seen in Table 10, Model (3), with the lowest AIC value of 940.9, is the most desirable. Model (1), with an AIC of 1103.4, is the least favorable. These results are in line with the values of log-likelihood scores.

However, while AIC is insightful, it is not the sole determinant of a model's quality. Taking these measures together, all three models are statistically significant and well-fitted. However, Model (3) could be considered the preferable choice due to its lowest AIC and highest log-likelihood values, suggesting a superior balance of fit and simplicity while still robustly rejecting the null hypothesis, H_0 .

Even though interpretation of pseudo R-squared is not as straightforward as it would be for linear models, it is introduced briefly. Statistical software Python offers, as a default, McFadden's R-squared when executing logistic regression. Models (2) and (3) display pseudo R-Squared values of 0.11 and 0.12, while Model (1) has a value of 0.1. However, unlike the R-squared in linear regression, this does not imply that the model explains 10% of the variance in the outcome variable. Instead, performance metrics discussed in the upcoming section, such as accuracy and recall, should offer more useful insights into how well the models perform.

On an ending note for this chapter, the statistical significance of each variable among the models is analyzed: as the only change among the three models is the threshold shift of the crash, it is noteworthy that the statistical significance deteriorates for TED spread, Yield curve and Commodity index, when threshold shifts from the original 4%, either to 5% or to 3%. Conversely, the most robust variables initially observed, BSEYD, Skew and OIX, either retain or increase the statistical significance when the threshold is adjusted. This observation underlines that these three variables are particularly resistant to changes in the crash threshold, thus demonstrating their consistent reliability in the predictive models, regardless of the defined crash criteria.

5.2.1 Accuracy metrics

Preceding the introduction of accuracy metrics, discussing the process involved in deriving these measures is imperative. For each observation in a logistic regression model, the logistic function values are computed. This function, as described in equation (12), yields a probability range from 0 to 1, utilizing the corresponding variable values, X_k, at a specific time (t) and their respective coefficients, β_k, from Table 10 as inputs.

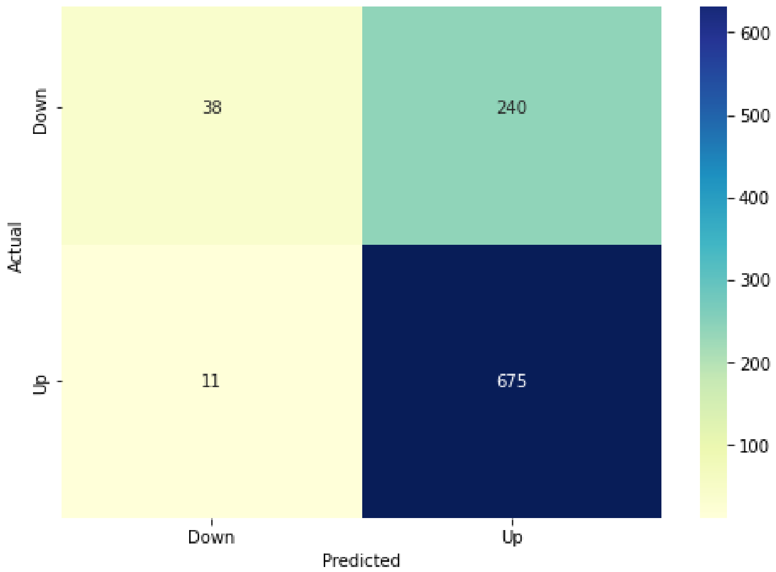
A threshold of 0.375 is subsequently established based on the logistic function's computed values. Determined by calculation (16), this threshold is aimed at optimally balancing the true positive rate and false positive rate using the Receiver Operating Characteristic (ROC) curve, which is introduced later. Notably, the threshold-setting process is carried out with out-of-sample data. As a result, logistic function values less than 0.375 are classified as '0', indicating an imminent market crash at time t+1. On the contrary, values exceeding the threshold are classified as '1', suggesting that a market crash is not anticipated at time t+1, and that the market is expected to appreciate.

$$Optimal\ treshold = MIN(\sqrt{(1 - tpr)^2 + fpr}\,), \tag{16}$$

While a threshold is set at 0.5 by default, this study made an attempt to enhance the accuracy of negative predictions by lowering the threshold. As indicated by the calculation, this approach allows our regression model to predict fewer number of crashes but with increased accuracy.

Figure 15 has the results for model (2) graphed with the help of the confusion matrix. Confusion matrix has two dimensions for actual and predicted values, and it is widely use tool for classification problems. For the full sample, the regression estimated 38 crash observations correctly, while it predicted 11 crashes incorrectly. Non-crash events were predicted correctly 675 times and incorrectly 240 times.

Figure 15. Confusion matrix, 'Crash4Pct', full sample



Six distinct calculations are presented for a comprehensive understanding of the confusion matrix. True Positives (*TP*) represent instances where the market was correctly predicted to rise and can be located in the bottom-right corner of the confusion matrix. True Negatives (*TN*) represent situations where the market downturn was correctly forecasted and can be found in the top-left corner. On the other hand, false negatives *(FN)* denote cases where the market rose contrary to the predicted downturn and are located at the bottom-left. Conversely, false Positives (*FP*) are instances where an upward movement was predicted, but the market fell instead. *FP*s can be located in the top right corner of the confusion matrix.

$$Accuracy = \frac{(TP+FN)}{(TP+FP+FN+TN)}, \quad (17)$$

$$Precision = TP \,/\, (TP + FP), \quad (18)$$

$$Recall = TP \,/\, (TP + FN), \quad (19)$$

$$F1\,Score = 2 * (precision * recall) \,/\, (precision + recall), \quad (20)$$

While typically only calculations from (16) to (20) are employed, in this study, negative predictions (21) and (22) are of particular interest. Thus, negative precision and true negative rate, or negative recall, are introduced.

$$Negative\ Precision\ =\ TN\ /\ (TN + FN), \tag{21}$$

$$Negative\ Recall\ =\ TN - (TN\ +\ FP), \tag{22}$$

Table 11 contains the results for full sample performance utilizing the abovementioned calculations. The corresponding confusion matrices for models 1 and 3 are in the appendix.

Table 11. Full-sample performance, accuracy

	Crash3Pct (1)	*Crash4Pct (2)*	*Crash5Pct (3)*
Accuracy	0.68	0.71	0.77
Precision	0.71	0.74	0.79
Recall	0.98	0.98	0.98
F1 Score	0.82	0.84	0.88
Precision negative	0.78	0.78	0.75
Negative recall	0.16	0.14	0.16

The general trend gauged from the table 11 is that the higher the threshold for the crash, the better the model tends to perform, especially in predicting market appreciation. A convenient method for understanding the performance could be examining the accuracy or the F1 score, which computes the harmonic mean of precision and recall, thereby providing a comprehensive single metric reflecting the model's overall performance.

As previously stated, Model (1) represents a crash threshold of 3%, while Model (3) applies a threshold of 5%. Notably, Model (1) correctly classified 68% of all market movements as either ups or downs, whereas Model (3) demonstrated superior accuracy, correctly predicting 77% of movements, being significantly more accurate. This does not, however, translate to increased accuracy in predicting market crashes. Interestingly, Model (1) outperforms Model (3) and (2) in predicting negative market instances correctly, demonstrating its superior performance in predicting market downturns.

The negative recall, which stands for correctly predicted negative instances out of all negative instances, is relatively low at 16% for model (1), which means that even though the model is accurate in its’ predictions of market downturns, it does not predict them frequently. This could be explained due to imbalances in the dataset, which contain more "1" observations than "0s". Such a scenario might lead the model to overpredict "1" outcomes, thus producing more false positives and eventually resulting in a lower negative recall. This hypothesis is further supported by the models’ extremely high positive recall score of 98%.

Negative Precision denotes the proportion of correctly predicted negative instances out of all instances the model predicted as negative. In this case, when model (1) predicts that a crash event will occur, it is correct 78% of the time. This good score indicates that the model accurately predicts negative instances.

In equity markets, achieving over 50% accuracy in precision or negative precision could carry substantial economic significance. This is simply because making accurate predictions more than half the time can lead to profitable trading strategies, thereby facilitating a competitive advantage to the proprietor of such a model.

Viewing performance metrics in absolute terms rather than percentages can offer a unique perspective, particularly as negative precision remained highly similar across the models. Table 12 offers a view of the absolute predicted values of the three models, capturing aspects that may not be apparent from percentage metrics alone.

Table 12. Full sample performance, absolute values

	Crash3Pct (1)	*Crash4Pct (2)*	*Crash5Pct (3)*
True Positives	641	675	726
False Positives	260	240	190
True Negatives	49	38	36
False Negatives	14	11	12

Table 12 discloses that no significant difference exists in the absolute value of true negatives between models (2) and (3). On the other hand, model (1) displays as many as 13 more correctly predicted negatives than model (3). As anticipated, model (3) exhibits the highest number of true positives, attributable to its highest crash detection threshold.

Figure 16 portrays the plot of the logistic function for the 4% threshold and the S&P 500 time series. The figure indicates predicted crash observations as vertical red lines. The lower part of the figure exhibits logistic function values ranging between 0 and 1, with the threshold of 0.375 represented by a black horizontal line. An examination of the figure reveals that the model predicted the crashes during the pandemic period with considerable reliability. However, between observations 400 and 700, several crash instances were not identified by the model, which could warrant the conclusion that there was a remarkably strong bull market during this period. Notably, even with the lower threshold of 3%, only a few instances during this upward-trending market period were accurately flagged as crashes (refer to Appendix 6 for details). In this period, the function fails to plunge to adequately low values.

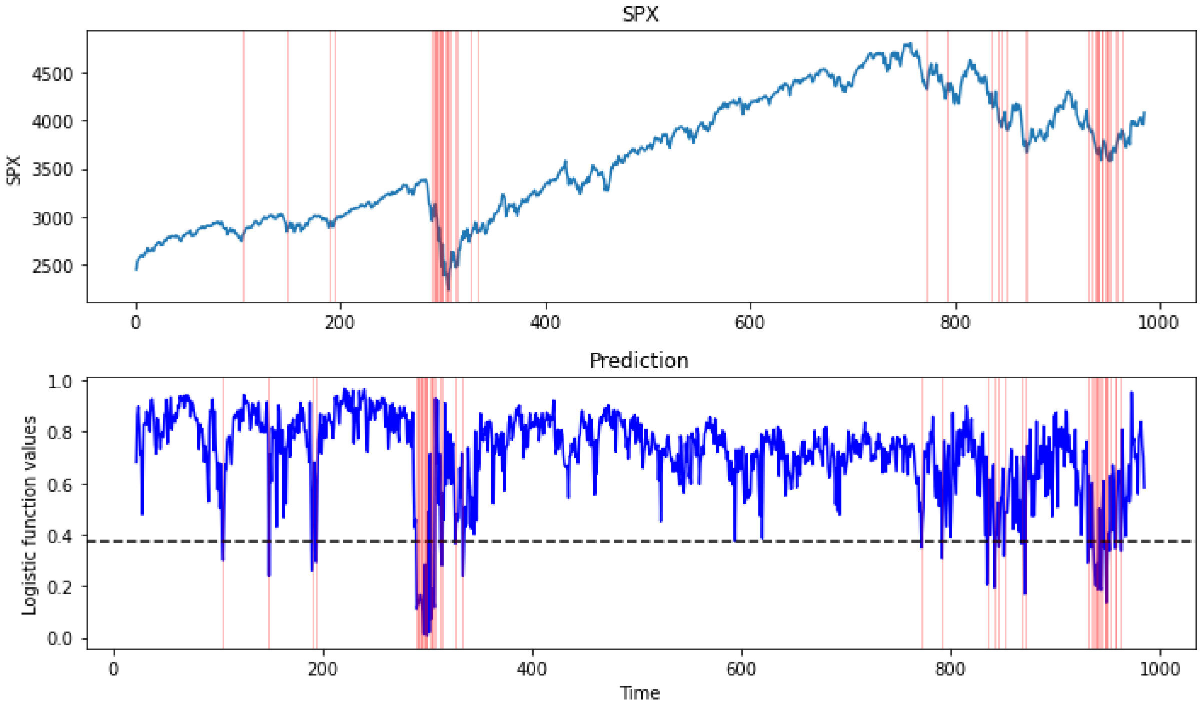


Figure 16. Full sample logistic function 'Crash4Pct' and S&P 500 time series

5.3 Performance analysis: out-of-sample regression models

Subsequent to the full sample performance analysis, the dataset is divided into two sets: the training set, which comprises the initial 85%, or 819 observations, of the data, which trains the model, and the testing set, encompassing the remaining 15% or 145 observations, which is the sample used to test the model performance with unseen data. Evaluating the model's out-of-sample performance is crucial as it is a more rigorous test and offers valuable insights into its potential real-world applicability.

As evidenced by the confusion matrix in Figure 17, the out-of-sample test results present a considerably more balanced distribution of outcomes than the in-sample results. The model successfully identified 34 instances as true negatives and 58 instances as true positives. These figures demonstrate the model's capability to predict both positive and negative outcomes accurately.

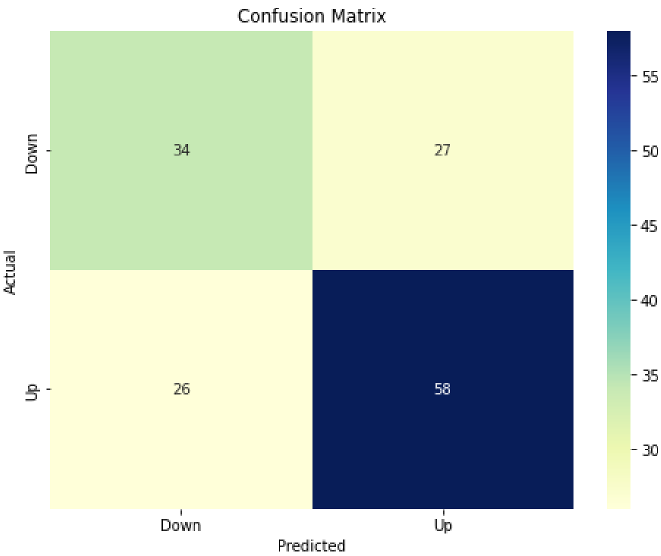


Figure 17. Confusion matrix, 'Crash4Pct', out-of-sample

However, it is worth noting that the model also produced 27 false positives and 26 false negatives. While these numbers are not insignificant, they are fairly balanced, suggesting that the model does not display a bias towards overpredicting either positive or negative outcomes, as may have somewhat been the case with the full sample data.

Table 13 below displays the accuracy metrics and absolute values for all three models computed from the out-of-sample data. A notable observation is the significant increase in the true negative score, also known as negative recall, compared to the full, in-sample results. This suggests an increased capacity of the model to identify instances of the negative class correctly. Furthermore, after the expected decline in accuracy, the negative precision for models (1) and (2) remains above the 50% mark, suggesting that these models retain their information edge over a random guess. This suggests that the models' predictive capacities are robust and adaptive.

One plausible explanation for the enhanced true negative score could be the more balanced nature (i.e., it contains a more even number of 1 and 0s) of the out-of-sample dataset compared to the full sample set. Specifically, the out-of-sample have

((34+27)/145) 42% of zero observations and 58% of 1s. The balance is considerably better than a full sample balance of 28% and 72%, respectively. This is due to 2022 being a negative year for equities, which led to increased "0" observations compared to, for example, the year 2021, which was a strong year for equities.

Table 13. Out-of-sample performance, accuracy metrics with absolute values

	Crash3Pct (1)	*Crash4Pct (2)*	*Crash5Pct (3)*
True Positives	53	34	51
False Positives	26	27	24
True Negatives	39	34	33
False Negatives	27	26	37
Accuracy	55%	58%	61%
Precision	67%	68%	68%
Recall	66%	69%	58%
F1 Score	67%	69%	63%
Precision negative	59%	57%	47%
Negative recall	60%	56%	58%

Both precision and F1-score have exhibited a decline, however, not to a degree that raises concern. Notably, the F1 score remains robust at 69% for model (2), indicating that the model maintains a fine balance between precision and recall. The most concerning insight from Table 13 is that model (3) no longer has the information edge predicting crash observations, as its negative precision has fallen to 47%. In addition, the table shows that with the out-of-sample data, model (1) predicts crash observations most successfully, with an accuracy of 59% and a true negative score of 60%, firming its ground as the most reliable model for predicting market downturns.

The observed outcomes could potentially be attributed to the limited number of crash observations in the training set for model (3). Originally, in the full-sample test, it had 226 '0' observations, as shown in Table 1. Accounting for the difference between total '0' observations and '0' observations in the out-of-sample test, model (3) only had (226-

57) 169 crash observations for training. In contrast, model (1) had a considerably larger pool of (309-61) 248 crash observations for training.

While certain statistical indicators such as log-likelihood and Chi-square suggested a superior performance of the model (3), the actual predictive performance, as observed in both in-sample and out-of-sample tests, contradicts this. This discrepancy suggests that despite model (3)'s superior statistical metrics, indicating lower error variance and better model fit, it does not perform as effectively in predicting actual market crashes. Therefore, the 3% threshold, which might be subject to a greater degree of volatility or noise, can still provide critical and actionable insights and prove more effective for real-world market downturn prediction.

Figure 18 displays the logistic function output for the out-of-sample data and the S&P 500 time series, most recent 145 observations, covering the period from May 5, 2022, to November 30, 2022. The graph visually represents the model's predictions with the market performance during this period. Similarly, as in Figure 16, red vertical lines denote predicted crash observations.

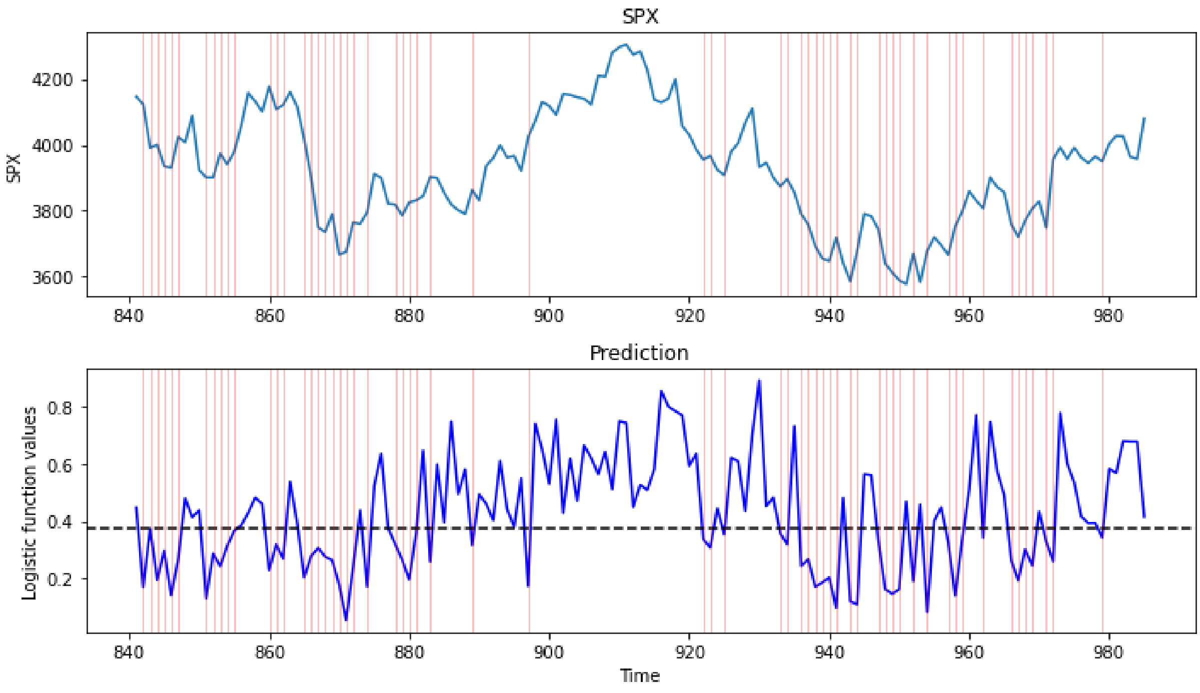


Figure 18 Out-of-sample logistic function ‘Crash4Pct’ and S&P 500 time series

In the concluding stage of the out-of-sample model evaluation, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are presented in Figure 19 The ROC curve is a tool for understanding the model's performance, showing model performance at all threshold levels. By employing the ROC curve and calculation (16), the thresholds for models (1), (2), and (3) were determined to be 0.39, 0.375, and 0.45, respectively. The thresholds were not altered between full-sample and out-of-sample results.

The AUC serves as a summary metric of the model's ability to distinguish between positive and negative classes. In this instance, the model (2) achieved an AUC of 0.63. This score, exceeding the baseline of 0.5 (representing a model with no discrimination ability), suggests a good fit for the model even with the out-of-sample data. In other words, the model demonstrates a reasonable capability to differentiate between market rise and fall periods. An AUC score closer to 1 would indicate a perfect model. However, such perfection is unlikely to be attained in the context of stock market predictions, at least with out-of-sample data.

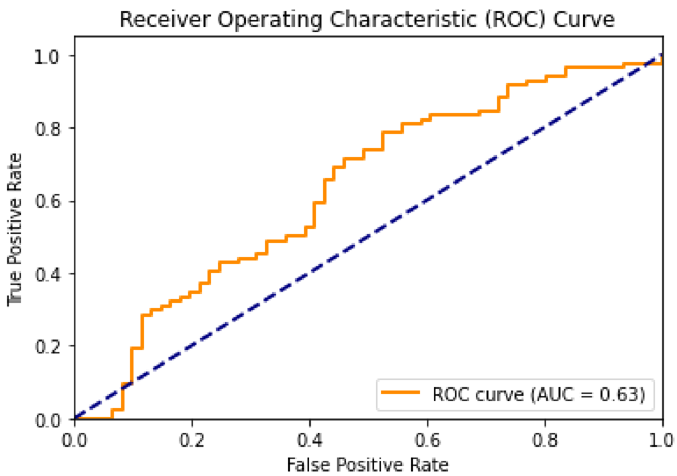


Figure 19. Receiver operating characteristic curve, 'Crash4Pct', out-of-sample

Further adding robustness to the study, models (1) and (3) too surpassed the AUC baseline of 0.5, scoring AUC values of 0.65 and 0.58, respectively. As with other performance metrics, model (1) presents the most promising results among the three, thereby suggesting a higher predictive capability in the given context.

6 Conclusions

This thesis aimed to elucidate the intriguing, challenging and extensively researched subject of forecasting future negative equity returns through constructing a composite indicator for stock market corrections. If successful, this tool would prove invaluable for anyone seeking an informational edge in the dynamic environment of the financial markets. Central to this thesis was the empirical examination of the efficient market hypothesis in the S&P 500 index from 2019 through 2022, during which the existence of market inefficiencies was observed.

The study employed a logistic regression model to forecast the likelihood of negative market movement at time *t+1,* with publicly available information at time *t.* An extensive literature review guided the selection of a composite of macroeconomic, financial, and option metric indicators to serve as predictive variables for the regression model. Among the eight chosen predictors, Volatility Smirk, Open Interest Difference, and Bond-Stock Earnings Yield Differential (BSEYD) emerged as statistically significant predictors of stock market corrections, with their statistical significance being notable at the 1% level, and thus also satisfying the higher t-statistic requirement introduced by Harvey et al. (2016).

Derived from the Fed model, BSEYD further affirmed its robustness as an invaluable component of the correction prediction models. It exhibited powerful predicting capabilities even when the market was affected by adverse events such as the COVID-19 pandemic crash, the bull market that followed, the War in Ukraine, and the historically high inflation period in 2022.

Blacks (1975) finding that information is first revealed in the options markets is supported by the results, as evidence of the still-existing lead-lag relationship between the option and equity market is found. The volatility smirk is a testament to this enduring phenomenon, consistently demonstrating robust predictive capabilities from Bates's 1991 study to that of Xing et al. in 2010. Despite the increasingly sophisticated models being developed by market participants to predict market returns and volatility, these informed views are often first manifested in the options market. As a result, a metric such as the open interest difference maintains its ability to track informed investors' 'smart money', yielding valuable and potentially profitable insights into market dynamics.

Contrasting previous research, the TED Spread showed only limited statistical significance, while the Commodity Index, Term Spread, Baltic Dry Index, and Option Bid-Offer Spread did not demonstrate substantial predictive power in anticipating equity market corrections. It was hypothesized that the variables lost most of their predictive power when first differenced to meet the logistic regression assumptions. Nevertheless, three of the former variables were kept in the final models, as the data input they provide to the model is unique, and their general capability of reflecting the credit risk, economic environment, or inflation cannot be overlooked.

In this research, the market correction served as the dependent variable, being assigned a '0' during a crash and '1' otherwise. Three distinct models were introduced to enhance the robustness of the findings, each differing in their definition of a market crash. Models 1,2 and 3 were assigned a '0' for each observation in the drawdown when the market correction exceeded 3%, 4% and 5%, respectively. All regression models robustly rejected the null hypothesis, H_o: "Predictors at time 't', have no effect on the likelihood of a crash at time 't+1' ", at 1% level.

The models were initially trained and evaluated using the full, in-sample dataset, yielding valuable insights into their predictive efficacy. Model (3) exhibited superior statistical results, displaying the most robust Akaike Information Criterion and log-likelihood, indicating best fit to the data. However, this statistical superiority did not convert to performance, as Model (1) was superior in correctly predicting the absolute number of negative instances. Nevertheless, all models displayed strong performance, achieving precision rates of over 75% in predicting market crashes. Even though models were accurate in their crash predictions, they only predicted them rarely, on average, in 16% of total instances. This low negative recall was hypothesized to stem from the dataset's imbalanced distribution of '0' and '1' instances which might have led the models to overpredict '1' instances and consequently increase the number of false positive predictions.

While the promising in-sample results indicated model efficacy, an out-of-sample assessment was crucial to validate performance under unseen conditions. Model (1) emerged as the most powerful, demonstrating a remarkable rise in crash prediction instances from 16% to 60%. Even though precision dropped from 78% to 59%, it still demonstrates robust ability to identify crashes on unseen data. Conversely, Model (3) lost its predictive power, with a negative precision of only 47%.

The area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve also indicated a satisfactory degree of fit for the models. Particularly for model (1), the out-of-sample AUC value of 0.65 demonstrates a commendable capability to distinguish between periods of market rise and fall. This indicator further strengthens the evidence of Model (1) superiority in predicting stock market corrections using out-of-sample data.

The study's findings and the constructed composite indicator assist investors with navigating and capitalizing on market downturns through informed decision-making in the financial markets. Additionally, the indicator’s predictive capabilities can be used as a risk management tool for market participants prioritizing capital preservation amidst market volatilities. Implementing this tool into risk assessment frameworks mitigates potential losses, through reducing volatility, enhancing portfolio resilience and optimizing long-term performance.

Future research could address the limitations identified in this study. The reduced predictive power observed in certain variables after first differencing could be combated with advanced statistical or machine learning methodologies that integrate these variables more effectively. Additionally, the dataset's imbalance towards '1' instances might have caused the model to overpredict market rises. Utilizing longer-term datasets with novel sampling methods could mitigate this bias, improving the model's accuracy. Finally, a longer-term data set would also add to the robustness of the findings by including a wider plethora of market events.

References

Bakshi, G., Panayotov, G., & Skoulakis, G. (2010). The Baltic Dry Index as a predictor of global stock returns, commodity returns, and global economic activity. Commodity Returns, and Global Economic Activity (October 1, 2010).

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. Handbook of the Economics of Finance, 1, 1053-1128.

Bates, D. S. (1991). The crash of '87: was it expected? The evidence from options markets. The journal of finance, 46(3), 1009-1044. https://www.jstor.org/stable/pdf/2328552.pdf?casa_token=JgMTRg6Hw7YAAAAA:CeqrxwB6aOB-tvjiVP3PYOoC6hBK8p6LsiFma-7XSO0rT-_Jw1a03Uao6xyl_R9_kZd3WmDh0XgBijurZS6qjglCed36igvIB61JKZQ7Naoex-ubeyViY

Berge, K., Consigli, G., & Ziemba, W. T. (2008). The Predictive Ability of the Bond-Stock Earnings Yield Differential Model. Journal of portfolio management, 34(3), 63-80. https://doi.org/10.3905/jpm.2008.706245

Black, F. (1975). Fact and Fantasy in the Use of Options. Financial analysts journal, 31(4), 36-72. https://doi.org/10.2469/faj.v31.n4.36'

Boudt, K., Paulus, E. C., & Rosenthal, D. W. (2017). Funding liquidity, market liquidity and TED spread: A two-regime model. Journal of empirical finance, 43(September), 143-158. https://doi.org/10.1016/j.jempfin.2017.06.002

Candelon, B., Piplack, J., & Straetmans, S. (2008). On measuring synchronization of bulls and bears: The case of East Asia. Journal of banking & finance, 32(6), 1022-1035. https://doi.org/10.1016/j.jbankfin.2007.08.003

Chen, J. (2022). Guide to Stock Market Crash. Investopedia. https://www.investopedia.com/terms/s/stock-market-crash.asp

Chen, S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. Journal of banking & finance, 33(2), 211-223. https://doi.org/10.1016/j.jbankfin.2008.07.013

Cheung, W., Fung, S., & Tsai, S. (2010). Global capital market interdependence and spill-over effect of credit risk: Evidence from the 2007-2009 global financial crisis. Applied financial economics, 20(1-2), 85-103. https://doi.org/10.1080/09603100903262962

Choi, H., & Jayaraman, N. (2009). Is reversal of large stock-price declines caused by overreaction or information asymmetry: Evidence from stock and option markets. The journal of futures markets, 29(4), 348-376. https://doi.org/10.1002/fut.20360

Cook, R. D. (2000). Detection of Influential Observation in Linear Regression. Technometrics, 42(1), 65-68. https://doi.org/10.1080/00401706.2000.10485981

DE LONG, J. B., SHLEIFER, A., SUMMERS, L. H., & WALDMANN, R. J. (1990). Positive Feedback Investment Strategies and Destabilizing Rational Speculation. The Journal of finance (New York), 45(2), 379-395. https://doi.org/10.1111/j.1540-6261.1990.tb03695.x

Easley, D., O'Hara, M., & Srinivas, P. (1998). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. The Journal of finance (New York), 53(2), 431-465. https://doi.org/10.1111/0022-1082.194060

Easley, D., O'Hara, M., & Srinivas, P. (1998). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. The Journal of finance (New York), 53(2), 431-465. https://doi.org/10.1111/0022-1082.194060

Estrella, A., & Trubin, M. (2006). The yield curve as a leading indicator: Some practical issues. Current issues in Economics and Finance, 12(5).

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of finance (New York), 25(2), 383. https://doi.org/10.2307/2325486

Fama, E. F. (1995). Random Walks in Stock Market Prices. Financial analysts journal, 51(1), 75-80. https://doi.org/10.2469/faj.v51.n1.1861

Fama, E. F., & French, K. R. (1988). Permanent and temporary components of stock prices. Journal of political Economy, 96(2), 246-273. https://www.jstor.org/stable/pdf/1833108.pdf?casa_token=3yBlLE-fOZkYAAAAA:igjGYV69l_9yeYky48T84uPSU2xJqM24RiZMZEVeiShELy-hfyrq34P3WMdvQuhjDhW20sKt29pYO1rcDnllb0tx8bKweBWw-DOm79bt2G3LCAkmGAOw

Federal Reserve Bank of New York, The Yield Curve as a Leading Indicator, https://www.newyorkfed.org/research/capital_markets/ycfaq.html

Fishburn, P. C. (1968). Utility theory. Management science, 14(5), 335-378. https://www.jstor.org/stable/pdf/2628674.pdf?casa_token=5VJ6EwNPR7IAAAAA:cCRVGmjwk6p8rc7VOlClzc87_pQw8hUWuEeR53BMwv2c5M6GMukKvZeSFYAM_9CdFMYqFbZS4J-JWZjy9OI6znBKJTkwbP0Bh9JYfpHoEFuPPpsNi5Xx

Fodor, A., Doran, J., & Krieger, K. (2011). Do option open-interest changes foreshadow future equity returns? Financial markets and portfolio management, 25(3), 265-280. https://doi.org/10.1007/s11408-011-0164-z

Gârleanu, N., Pedersen, L. H., & Poteshman, A. M. (2009). Demand-Based Option Pricing. The Review of financial studies, 22(10), 4259-4299. https://doi.org/10.1093/rfs/hhp005

Glosten, L. R., & Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of financial economics, 14(1), 71-100. https://doi.org/10.1016/0304-405X(85)90044-3

Harvey, C. R. (1989). Forecasts of Economic Growth from the Bond and Stock Markets. Financial analysts journal, 45(5), 38-45. https://doi.org/10.2469/faj.v45.n5.38

Harvey, C. R., Liu, Y., & Zhu, H. (2016). and the Cross-Section of Expected Returns. The Review of financial studies, 29(1), 5-68. https://doi.org/10.1093/rfs/hhv059

Hertzberg, M. P., & Beckman, B. A. (1989). Business cycle indicators: Revised composite indexes. Survey of current business, 69(1), 23-28.

HESTON, S. L. (1993). Invisible Parameters in Option Prices. The Journal of finance (New York), 48(3), 933-947. https://doi.org/10.1111/j.1540-6261.1993.tb04025.x

Hjalmarsson, E. (2010). Predicting Global Stock Returns. Journal of financial and quantitative analysis, 45(1), 49-80. https://doi.org/10.1017/S0022109009990469

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R. Springer. https://hastie.su.domains/ISLR2/ISLRv2_website.pdf

Kaeck, A., van Kervel, V., & Seeger, N. J. (2022). Price impact versus bid–ask spreads in the index option market. Journal of financial markets (Amsterdam, Netherlands), 59, 100675. https://doi.org/10.1016/j.finmar.2021.100675

Lewellen, J. (2004). Predicting returns with financial ratios. Journal of financial economics, 74(2), 209-235. https://doi.org/10.1016/j.jfineco.2002.11.002

Liang, C., Ma, F., Li, Z., & Li, Y. (2020). Which types of commodity price information are more useful for predicting US stock market volatility? Economic modelling, 93, 642-650. https://doi.org/10.1016/j.econmod.2020.03.022

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. The journal of finance, 20(4), 587-615.

Lleo, S., & Ziemba, W. T. (2017). Does the bond-stock earnings yield differential model predict equity market corrections better than high P/E models? Financial markets, institutions & instruments, 26(2), 61-123. https://doi.org/10.1111/fmii.12080

Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. The Review of financial studies, 1(1), 41-66. https://doi.org/10.1093/rfs/1.1.41

Lo, A. W., & MacKinlay, A. C. (1999). A non-random walk down Wall Street. In A Non-Random Walk Down Wall Street. Princeton University Press. https://ebookcentral-proquest-com.proxy.uwasa.fi/lib/tritonia-ebooks/reader.action?docID=819621

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. The Journal of finance (New York), 55(4), 1705-1765. https://doi.org/10.1111/0022-1082.00265

Lyócsa, Š., Baumöhl, E., & Výrost, T. (2022). YOLO trading: Riding with the herd during the GameStop episode. Finance research letters, 46, 102359. https://doi.org/10.1016/j.frl.2021.102359

Maio, P. (2013). The "Fed Model" and the Predictability of Stock Returns. Review of Finance, 17(4), 1489-1533. https://doi.org/10.1093/rof/rfs025

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), 59-82. https://pubs.aeaweb.org/doi/pdf/10.1257%2F089533003321164958

Mensi, W., Beljid, M., Boubaker, A., & Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. Economic modelling, 32, 15-22. https://doi.org/10.1016/j.econmod.2013.01.023

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. The Journal of finance, 32(4), 1151-1168.

Modigliani, F., & Cohn, R. A. (1979). Inflation, rational valuation and the market. Financial AnalystsJournal, 35(2),24-44. https://www.jstor.org/stable/pdf/4478223.pdf?casa_token=WaD1qAuKu68AAAAA:QzM_-wDbbY-cDKi4qd5Zqa5P2Hd6AIvdATE7BgJPN-hgHC6sC-CePjkkRFDyatP4dE0bkyVY28fzvp5dC35RK60XOgr-JVDt8y_izayFC8KHSvj6gVPA

Ofek, E., Richardson, M., & Whitelaw, R. F. (2004). Limited arbitrage and short sales restrictions: Evidence from the options markets. Journal of financial economics, 74(2), 305-342. https://doi.org/10.1016/j.jfineco.2003.05.008

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. Journal of financial economics, 63(1), 3-50. https://doi.org/10.1016/S0304-405X(01)00088-5

Papailias, F., Thomakos, D. D., & Liu, J. (2017). The Baltic Dry Index: cyclicalities, forecasting and hedging strategies. Empirical Economics, 52, 255-282.

Pesaran, M. H. (2010). Predictability of asset returns and the efficient market hypothesis (No. 3116). CESifo.

Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. International journal of forecasting, 21(1), 137-166. https://doi.org/10.1016/j.ijforecast.2004.05.004

S&P Global (2023). S&P GSCI Methodology. S&P Dow Jones Indices. https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-gsci.pdf

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. Industrial Management Review Spring 6, 51-49. https://investmenttheory.org/uploads/3/4/8/2/34825752/paulsamuelson-proof.pdf

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The journal of finance, 19(3), 425-442.

Shiller, R. J. (2000). Irrational exuberance. In Irrational exuberance. Princeton university press. https://doi.org/10.1515/9781400865536

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. The Journal of economic perspectives, 17(1), 83-104. https://doi.org/10.1257/089533003321164967

Shiller, R. J. (2015). Irrational exuberance. In Irrational exuberance. Princeton university press.

Sornette, D. (2003). Critical market crashes. Physics reports, 378(1), 1-98. https://doi.org/10.1016/S0370-1573(02)00634-8

Sornette, D. (2017). Why stock markets crash : Critical events in complex financial systems. Princeton University Press. https://ebookcentral-proquest-com.proxy.uwasa.fi/lib/tritonia-ebooks/detail.action?docID=4910404#

Stephanie Glen. (2023). "Cook's Distance / Cook's D: Definition, Interpretation" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/cooks-distance/

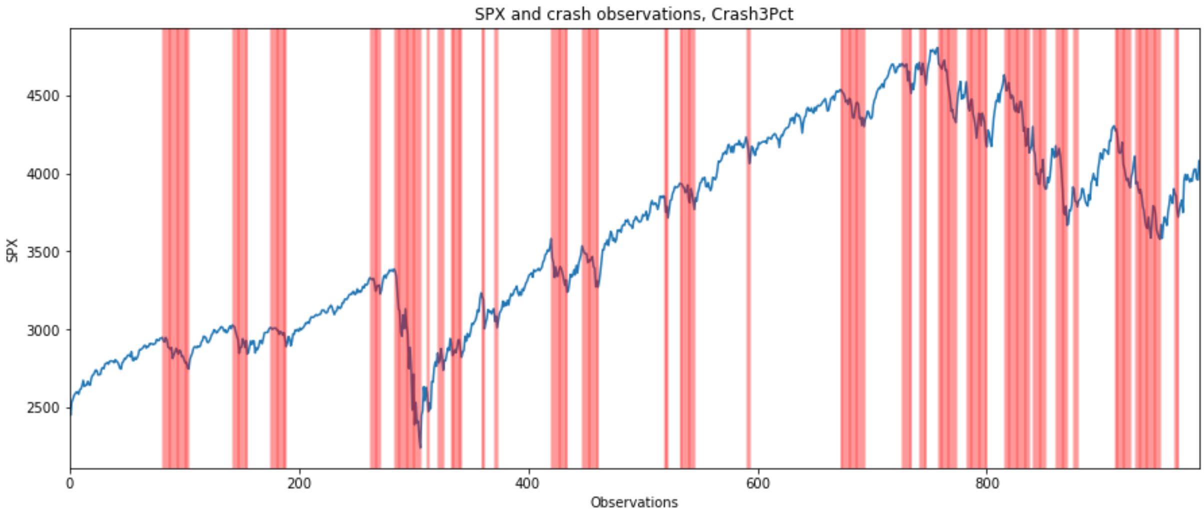
Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med. 2011 Oct;18(10):1099-104. doi: 10.1111/j.1553-2712.2011.01185.x. PMID: 21996075.

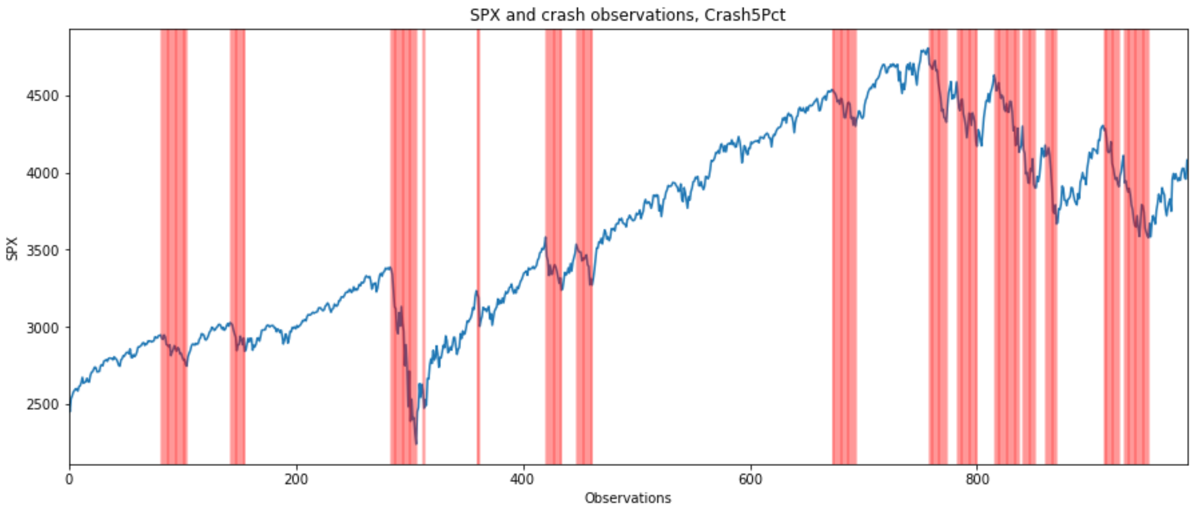
Sullivan, R., Timmermann, A., & White, H. (1999). Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. The Journal of finance (New York), 54(5), 1647-1691. https://doi.org/10.1111/0022-1082.00163

Xing, Y., Zhang, X., & Zhao, R. (2010). What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? Journal of financial and quantitative analysis, 45(3), 641-662. https://doi.org/10.1017/S0022109010000220

Appendix

Appendix 1. Crash observations with 3% and 5% thresholds

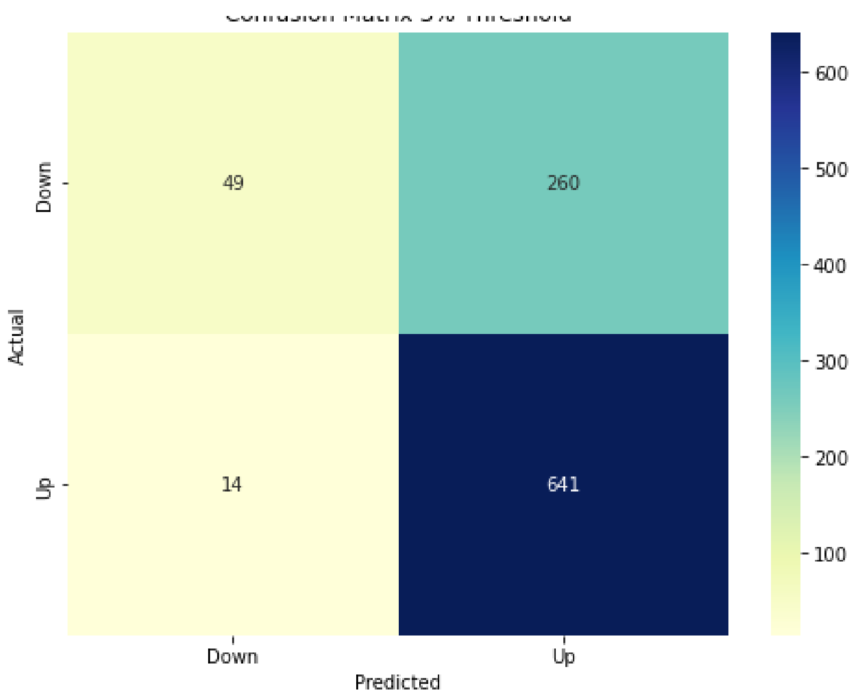


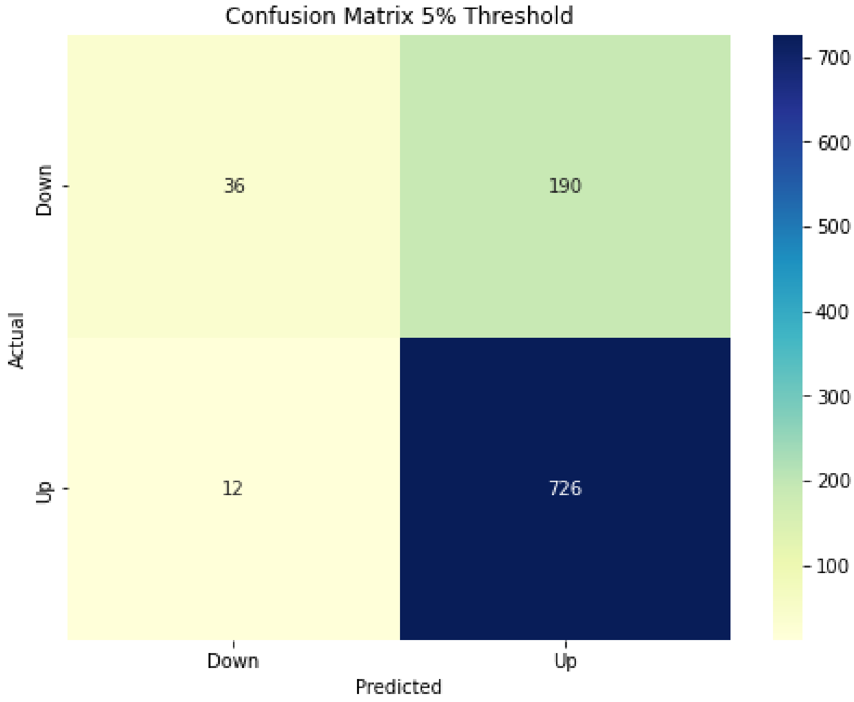


Appendix 2. Detailed overview of drawdown statistics

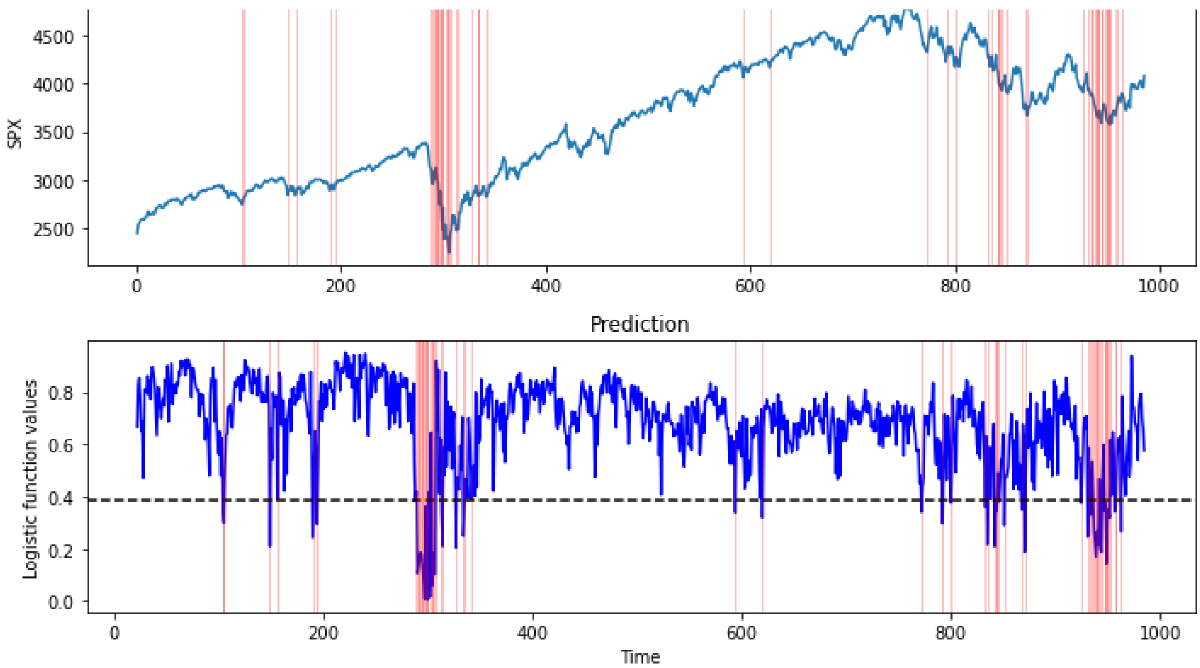
Threshold:	3%		4%		5%	
#	drawdown (%)	duration (days)	drawdown (%)	duration (days)	drawdown	duration
1	6.8%	34	6.8%	34	6.8%	34
2	6.1%	19	6.1%	19	6.1%	19
3	4.1%	20	4.1%	20	33.9%	33
4	3.1%	14	33.9%	33	5.9%	2
5	33.9%	33	5.9%	2	7.1%	3
6	5.9%	2	4.1%	14	9.6%	21
7	3.8%	7	7.1%	3	7.5%	18
8	4.1%	14	9.6%	21	5.2%	32
9	7.1%	3	7.5%	18	10.0%	23
10	3.9%	3	4.2%	20	9.1%	27
11	9.6%	21	4.0%	5	10.8%	31
12	7.5%	18	5.2%	32	9.3%	15
13	3.7%	4	4.1%	13	12.2%	14
14	4.2%	20	10.0%	23	9.2%	21
15	4.0%	5	9.1%	27	13.0%	30
16	5.2%	32	10.8%	31		
17	4.1%	13	9.3%	15		
18	3.1%	10	12.2%	14		
19	10.0%	23	9.2%	21		
20	9.1%	27	13.0%	30		
21	10.8%	31	4.6%	6		
22	9.3%	15				
23	12.2%	14				
24	3.2%	6				
25	9.2%	21				
26	13.0%	30				
27	4.6%	6				

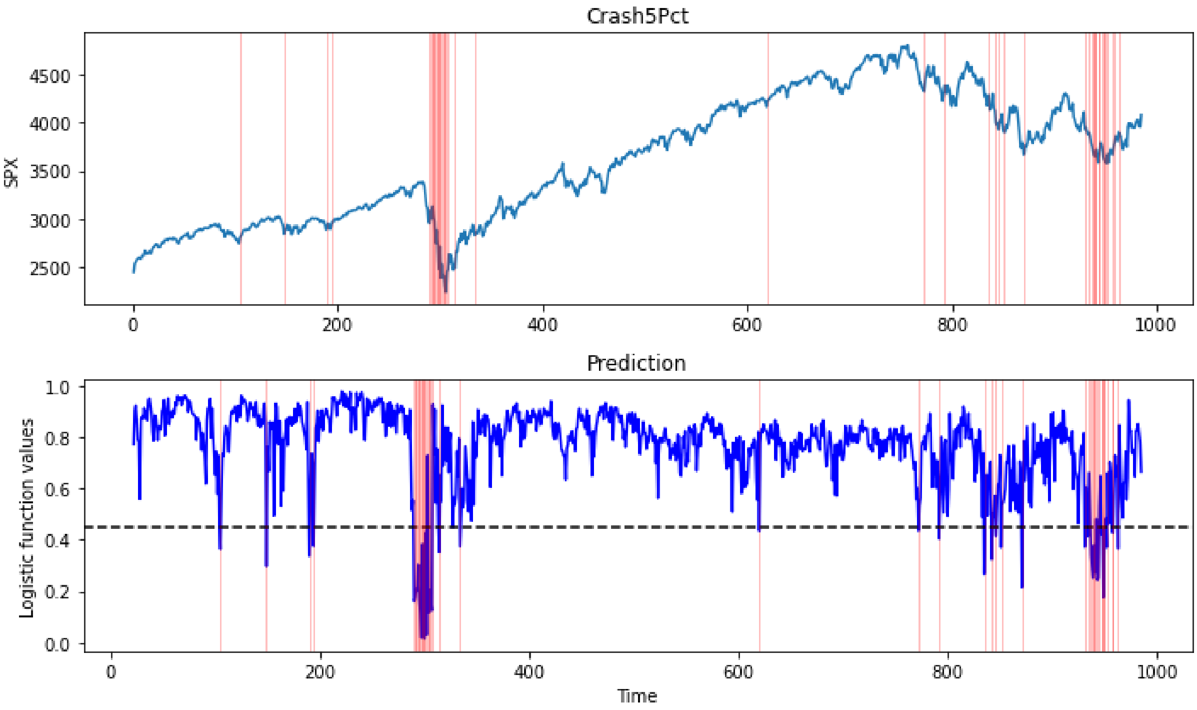
Appendix 3. Confusion matrices, alternative thresholds, full sample



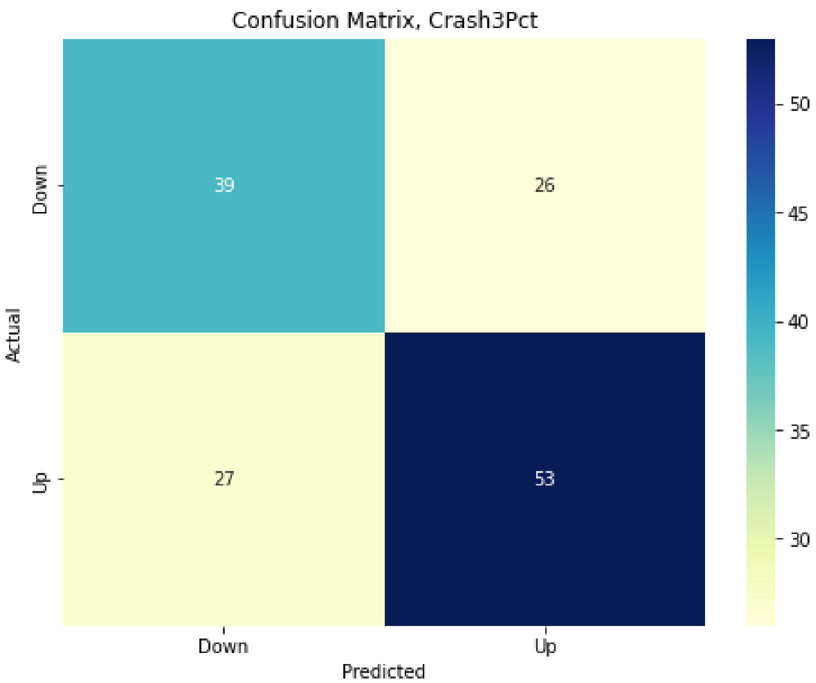


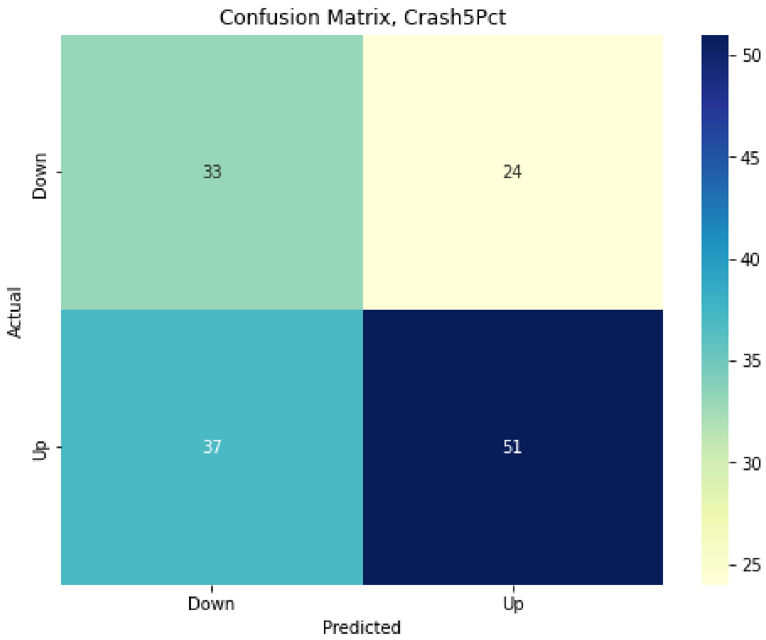
Appendix 4. Predicted crash observations, alternative thresholds, full sample



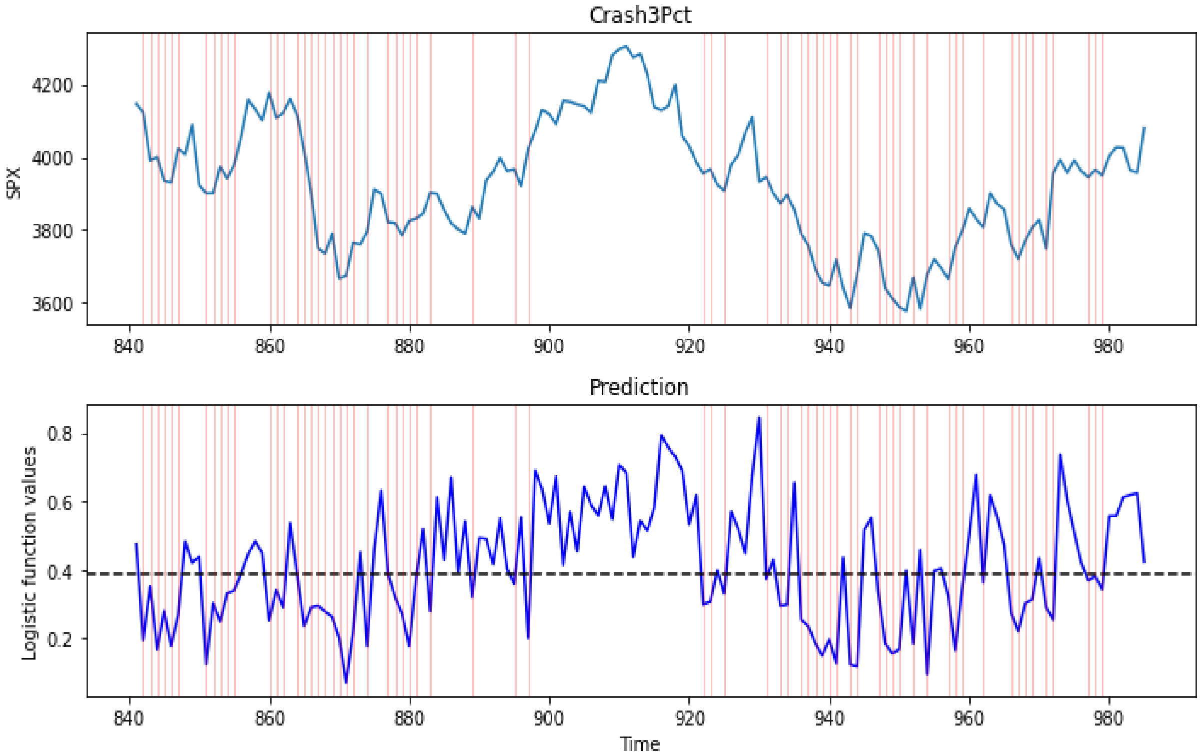


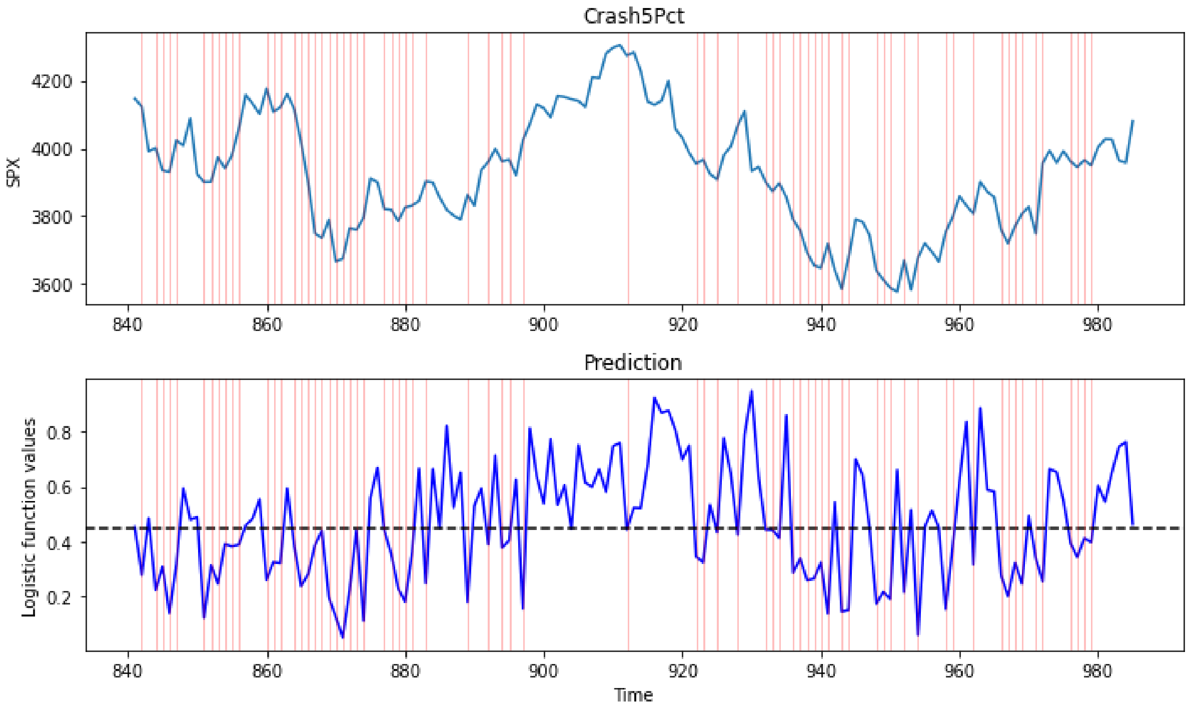
Appendix 5. Confusion matrices, alternative thresholds, out-of-sample




Confusion Matrix, Crash5Pct
Actual
Down
Up
33
24
37
51
Down
Up
Predicted
50
45
40
35
30
25

Appendix 6. Predicted crash observations, alternative thresholds, out-of-sample





Appendix 7. ROC curves, alternative thresholds, out-of-sample