Don't Stop Me Now! Identification and Prediction of Unnecessary Volatility Interruptions

Benjamin Clapham, Florian Ewald, Peter Gomber, Niklas Trimpe

Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

Abstract

Exchange operators use circuit breakers like volatility interruptions to prevent transitory or error-induced price shocks. However, these mechanisms can impair price efficiency and market quality when triggered by legitimate price movements due to new information. We introduce a clustering approach to identify unnecessary volatility interruptions that occur within persistent price trends, delaying price discovery without improving market quality. These interruptions are more likely to occur when liquidity is high, relevant news is present, and prices are already near the predefined triggering threshold. To improve market design, we propose a deep learning model that predicts unnecessary interruptions based on pre-interruption data.

Keywords: Circuit Breakers, Volatility Interruptions, Market Quality, Price Discovery, Market Design, Machine Learning

JEL: G14, G18

[⋆]This work was supported by efl – the Data Science Institute, Frankfurt. Deutsche Börse AG provided us with the data through their A7 analytics platform. Neither the financial support nor any other relationships or competing interests influenced the work reported in this paper.

[⋆⋆]We thank seminar participants at the 29th Forecasting Financial Markets Conference, the NYSE Microstructure Meets AI Conference, and efl – the Data Science Institute Jour Fixe at Goethe University Frankfurt for helpful comments.

[*]Corresponding author

Email address: clapham@wiwi.uni-frankfurt.de (Benjamin Clapham)

August 28, 2025

1. Introduction

Securities markets are frequently subject to sudden and unsubstantiated price fluctuations caused by liquidity shocks, erroneous orders, or unexpected firm-specific or macroeconomic news. The advent of algorithmic and high-frequency trading, which involves fully automated order submissions and cancellations executed at extremely high speeds (O'Hara, 2015), has amplified the frequency and severity of short-term liquidity imbalances and transient price dislocations (U.S. Securities and Exchange Commission, 2016). These heightened volatility risks inherent in modern securities markets have prompted both market operators and regulators to adopt mitigating measures. As a response, exchanges worldwide implement circuit breakers, which are market safeguards such as trading halts or volatility interruptions, that temporarily pause or slow trading during episodes of substantial price movement. These mechanisms are designed to promote price continuity and maintain the orderly operation of fully electronic securities markets by curbing transitory or error-induced price shocks.

Circuit breakers were first introduced on major stock exchanges in the late 1980s (e.g., Lee et al., 1994; Lauterbach and Ben-Zion, 1993) and have been increasingly adopted worldwide. In a 2016 survey among exchange operators (Gomber et al., 2016), 86% of the responding trading venues employ circuit breakers to ensure investor protection and market stability, up from 60% reported in a similar survey conducted by the World Federation of Exchanges in 2008 (World Federation of Exchanges, 2008). Moreover, regulators in most jurisdictions mandate the use of circuit breakers.[1] However, their effectiveness remains a subject of ongoing debate. Critics argue that circuit breakers unnecessarily disrupt the trading process, potentially hindering market efficiency by delaying price discovery (Fama, 1988) and causing volatility spillovers across time and markets (Subrahmanyam, 1994). Moreover, they are associated with adverse effects on market liquidity (Hautsch and Horvath, 2019).

The main issue with the design of current circuit breakers is their reliance on simplistic rule-based mechanisms, which trigger interruptions as soon as the price of a security or index crosses a pre-defined threshold. Thus, these mechanisms do not differentiate between legitimate market movements driven by fundamental information and erratic price jumps caused by liquidity shocks, erroneous orders, misconfigured trading algorithms, or fake news.[2] As a result, existing circuit breakers are characterized by a trade-off between their protective role in maintaining market stability and their potential adverse effects on market quality (Hautsch and Horvath, 2019).

The issue of simplistic circuit breaker rules leading to unnecessary trading interruptions becomes even more critical when considering the frequency of these events.

[1]For example, the U.S. Securities and Exchange Commission mandates a trading pause for individual stocks through the Limit Up/Limit Down mechanism in the event of sudden price swings, and a market-wide circuit breaker in the form of a trading halt in case of a severe decline in the S&P 500 Index (U.S. Securities and Exchange Commission, 2012). Similarly, the European financial market regulation MiFID II mandates trading venues to have mechanisms in place to halt or restrict trading in case of significant price movements (European Commission, 2014).

[2]See Bongaerts et al. (2024) for a related discussion on the issues arising from the simplistic and backward-looking nature of current circuit breaker designs.

Rather than being rare events, circuit breakers are frequently triggered, even in the most liquid stocks.[3] For instance, the European Securities and Markets Authority (ESMA) reported a peak of more than 3,000 circuit breaker events within a single week in March 2020 due to market turmoil related to COVID-19 (European Securities and Markets Authority, 2022). In our sample of German blue chip stocks, we observe an average of 5.27 interruptions per day and 161 interruptions per stock over the observation period of six years.[4] Consequently, identifying and avoiding unnecessary interruptions could significantly enhance the effectiveness of circuit breakers and improve overall market quality.

To address this issue and to contribute to solving the trade-off between market stability and the negative impact of unnecessary interruptions, this paper introduces a novel cluster-based approach utilizing single-stock volatility interruptions, the prevalent circuit breaker mechanism in Europe. This method allows for the identification of those volatility interruptions that are triggered during persistent price trends, where they merely delay the incorporation of new information into prices. To the best of our knowledge, this is the first framework designed to detect such unnecessary volatility interruptions. Our study is based on a dataset of 7,899 historical volatility interruptions in the 49 stocks that were included in the DAX40 benchmark index at any point during our observation period, which spans from April 2019 to December 2024. The dataset includes high-frequency order book, trade, and news data. By combining an autoencoder model that consolidates order book information before and after each interruption with a Gaussian Mixture Model (GMM), we classify 37% of the observed volatility interruptions as unnecessary, i.e., interruptions that delay the price discovery process and lead to a decrease in market quality.

To examine the drivers of unnecessary interruptions, an analysis utilizing probit regressions reveals that unnecessary volatility interruptions are more likely to occur during periods of high liquidity and the presence of relevant news. Economically, these conditions suggest that the price changes leading to the interruption are likely driven by shifts in value expectations and corresponding trading activity. In such cases, the rule-based nature of current circuit breaker mechanisms unnecessarily interferes with the trading process. Moreover, our results suggest that volatility interruptions are more likely to be unnecessary when the last price before the interruption is near the pre-defined threshold that triggers the circuit breaker. When prices approach these thresholds due to sustained movements throughout the trading day, even minor price fluctuations from normal trading activity can activate the circuit breaker, causing unwarranted market disruptions. While revealing the determinants of unnecessary volatility interruptions, the probit model is not able to predict unnecessary interruptions based on ex-ante order book, trade, and news information.

Based on these findings, we develop a deep learning model that significantly enhances prediction capabilities by effectively capturing complex spatial and temporal

[3] For an overview of the number of circuit breaker events in selected empirical studies, see Bongaerts et al. (2024), Table 2.

[4] After the cleaning steps shown in Table 2, the sample has an average of 4.59 interruptions per day and 140.55 interruptions per stock.

relationships among input features. Unlike probit regression models, it eliminates the need to aggregate input time series data and to pre-select relevant variables, as it is capable of processing raw data and autonomously identifying useful features. The model combines convolutional neural networks (CNNs), an inception module, and long short-term memory (LSTM) layers. The CNN component extracts features from individual order book snapshots, while the inception module focuses on time-wise convolutions, summarizing features over time. These features are then processed by the LSTM layer, capturing temporal changes essential for the classification task. Using this model, we are able to predict unnecessary volatility interruptions with a precision of 82.5% and a recall of 20.4%. The high precision ensures a low rate of false positives, i.e., avoiding interruptions only when they are truly unnecessary, while still capturing a meaningful share of all unnecessary interruptions. To assess the broader implications of deploying the model, we construct a simple welfare framework that compares the estimated gains and losses from implementing the prediction model alongside existing rule-based circuit breaker mechanisms. Our results show that the model generates a net welfare gain, which scales meaningfully with the number of stocks and trading venues where it is applied.

Our findings contribute to two strands of literature. First, our paper adds to the extensive research stream that discusses and analyzes the effectiveness of circuit breakers from both theoretical and empirical perspectives.[5] Circuit breakers can help to "cool down" markets and reduce volatility by providing market participants with time to reassess their trading strategies, inventories, and the impact of news (Ma et al., 1989). However, these safeguards can also disrupt trading and impair market liquidity. They may hinder inventory management by liquidity providers (Lauterbach and Ben-Zion, 1993), delay price discovery (Lehmann, 2019), and induce a "magnet effect", whereby prices accelerate toward the interruption threshold (Chen et al., 2024; Subrahmanyam, 1994).[6] Additionally, they can trigger volatility spillovers to other markets and future trading sessions (Subrahmanyam, 1994).

Despite these drawbacks, circuit breakers seem necessary to prevent erroneous price jumps in today's fully electronic securities markets (Subrahmanyam, 2013), where order submissions, executions, and price determination occur autonomously at millisecond frequency (O'Hara, 2015). Empirical studies reach contradictory conclusions regarding the effectiveness of circuit breakers in reducing volatility, although they generally agree on their harmful effects on liquidity and price discovery (e.g.,

[5]For an overview of the literature, refer to the surveys by Abad and Pascual (2013) and Sifat and Mohamad (2019).

[6]The magnet effect, which refers to the acceleration of prices toward a circuit breaker threshold, has been formalized and empirically validated by Chen et al. (2024). The authors demonstrate that as markets approach a trading halt, volatility increases sharply and returns exhibit increasingly negative skewness. These findings suggest that circuit breakers can distort trading behavior and tend to destabilize price formation during periods of significant market decline. Other empirical evidence on the magnet effect of circuit breakers or price limits is mixed. Some studies find support for its existence (e.g., Cho et al., 2003), while others do not (e.g., Abad and Pascual, 2007).

Abad and Pascual, 2010; Hautsch and Horvath, 2019; Kim and Rhee, 1997).[7] Based on an in-depth analysis of trading interruptions at Nasdaq, Hautsch and Horvath (2019) conclude that there is a trade-off between the protective role of trading interruptions and their potentially adverse effects on volatility, liquidity, and price efficiency. Variations in distributions across different observation periods and datasets, along with differences in the design of the safeguards (Clapham et al., 2017a), may explain the differing conclusions regarding their effectiveness in reducing volatility.

The study most closely related to ours is the model by Bongaerts et al. (2024), which demonstrates that properly calibrated circuit breakers can prevent market runs by curbing excessive trading. Like us, the authors argue that the current simplistic, price-triggered circuit breaker mechanisms fail to differentiate between legitimate liquidity demand and inefficient excessive trading. To enhance the current market design, they propose a forward-looking circuit breaker that becomes increasingly restrictive as the expected welfare losses from market runs increase. Another closely related study is Moise (2025), which examines the economic triggers of single-stock circuit breakers in the U.S. Although these mechanisms are intended to pause trading during high-volatility phases driven by informational shocks, the study finds that they are often triggered by short-term liquidity shocks and are followed by reduced liquidity across all types of stocks.

We contribute to the literature on circuit breakers by demonstrating that it is possible to identify and avoid unnecessary interruptions. This can mitigate the adverse effects of circuit breakers and potentially resolve the trade-off between their protective role and their negative impact on market quality.

Second, we contribute to the emerging body of literature that applies machine learning techniques to market microstructure research. For instance, Easley et al. (2021) utilize machine learning methods to predict future levels of liquidity, volatility, and other critical variables for market participants and researchers. Similarly, Kwan et al. (2021) employ reinforcement learning to investigate the price discovery process, while Sirignano and Cont (2019) leverage deep learning to forecast the direction of price movements based on historical limit order book (LOB) data. Our study extends the application of machine learning in market microstructure by exploring its use in analyzing circuit breakers and its potential to enhance their design and effectiveness.

Overall, our findings can enhance circuit breaker mechanisms by reducing the number of unnecessary interruptions. This can mitigate potential adverse effects of these mechanisms on market quality (Hautsch and Horvath, 2019; Subrahmanyam, 1994) and ultimately improve market efficiency. We also discuss the key considerations when implementing our proposed model in practice. Additionally, our findings can inform discussions with both market operators and regulatory authorities regarding the design and rules of circuit breakers.

[7] Exhibit 17.2 in the survey by Abad and Pascual (2013) provides a systematic comparison of empirical studies examining whether circuit breakers reduce volatility, improve price discovery, or interfere with liquidity and the trading process.

The remainder of the paper is organized as follows: Section 2 provides an overview of the institutional background, details on the dataset, and key descriptive statistics. Section 3 outlines our methodology for clustering volatility interruptions and presents the corresponding results. In Section 4, we describe a regression-based approach to examine economic drivers of unnecessary volatility interruptions and the implications for market quality. Section 5 introduces an advanced deep learning model for predicting unnecessary interruptions, evaluates its performance, and discusses effects on overall welfare. Section 6 presents robustness checks and discusses limitations, while Section 7 outlines key considerations for implementing our prediction model in practice. Finally, Section 8 concludes.

2. Institutional background & data

2.1. Volatility interruptions on Xetra

Our analyses are based on volatility interruptions, which are the common type of circuit breakers in European securities markets (Gomber et al., 2016). Instead of leading to a complete trading halt, volatility interruptions temporarily switch the trading phase from continuous trading to an unscheduled call auction in individual stocks once they are triggered. This is similar to the Limit Up/Limit Down mechanism in the U.S., suggesting that our results are transferable to other markets. Our sample of volatility interruptions represents data from the German trading venue Xetra. Xetra, operated by Deutsche Börse in Frankfurt, is a fully automated, order-driven trading system where buy and sell orders are matched based on price-time priority within a standard open LOB. The system provides continuous trading for the immediate execution of orders and scheduled auctions at specific times during the trading day. These auctions determine opening and closing prices by pooling liquidity and matching orders at a single clearing price. In contrast to other European markets, Xetra also has a scheduled auction in the middle of the trading day (hereafter referred to as intraday auction). Xetra is the most liquid market and reference market for German equities and ETFs.

Like most European trading venues, Xetra employs rule-based circuit breakers known as volatility interruptions to safeguard orderly price formation during continuous trading.[8] A volatility interruption is triggered at the individual stock level when the potential execution price of an incoming order breaches one of the predefined price thresholds. These thresholds are defined by two symmetric price corridors around reference prices: a dynamic range, centered on the last traded price, and a static range, centered on the most recent auction price. The volatility interruption mechanism is visualized in Figure 1.

[8]While our analysis focuses on the Single Volatility Interruption Model applicable to stocks, it is important to note that Deutsche Börse also uses the Automated Corridor Extension (ACE) model, introduced in November 2021, specifically for trading Exchange-Traded Funds (ETFs) and Exchange-Traded Products (ETPs). The ACE mechanism allows a sequence of progressively widening price corridors during periods of elevated volatility, aimed at facilitating a smooth transition back to continuous trading. As our study focuses solely on DAX40 stocks, the ACE mechanism does not apply to our dataset.

If the potential execution price falls outside either corridor, continuous trading is halted and trading switches to an unscheduled call auction, while the triggering order remains unexecuted. This auction phase shall help to rebalance the order book and ensure price continuity, thereby preventing disorderly trading conditions.[9]

The corridor widths are set individually by Deutsche Börse for each security, taking into account its historical volatility. While the exact corridor parameters are not publicly disclosed, the static corridor can be reasonably estimated by measuring the most extreme pre-interruption price movements relative to the most recent auction price. Both the dynamic and static price corridors remain constant throughout the trading day and are adjusted by the exchange operator only if a stock's historical volatility changes substantially, based on a weekly review process. By design, the dynamic corridor is narrower than the static one, allowing it to respond more sensitively to sudden price swings during continuous trading.

Each volatility interruption consists of a 2-minute call auction phase, which is extended by a random end between 0 and 30 seconds. This random end is designed to obscure the exact end time of the auction and thereby reduce the risk of strategic manipulation. If the final auction price remains outside an expanded version of the price corridor, the interruption may be extended and manually resolved by the exchange in consultation with the participant who triggered it (Deutsche Boerse Group, 2024).[10] Once the auction concludes, continuous trading resumes, and the price determined in the interruption auction becomes the new reference for the static corridor.

2.2. Data

Our dataset spans the period from April 2019[11] to the end of 2024. For the purpose of our analysis, we collect all instances of volatility interruptions affecting stocks included in the DAX40, Germany's benchmark stock index. While the DAX40 comprises the 40 largest listed German companies by market capitalization and is periodically revised, we consider the full set of 49 stocks (see also Figure 11 in Appendix A for a list of all included DAX constituents) that were constituents of the index at any point during the observation period. This approach ensures consistency by holding the stock universe constant throughout the analysis, and helps prevent survivorship bias by retaining all firms that were part of the DAX40 at any point during the observation period. We only partially exclude Wirecard AG from the sample starting on June 25, 2020, which marks the date of its insolvency filing, as it no longer qualifies as a highly liquid stock and its collapse to penny stock status could distort the dataset. Moreover, Linde PLC is included in the dataset only until February 27, 2023, the date of its complete delisting from Xetra.

[9] Although not the focus of this paper, volatility interruptions may also occur during scheduled auctions. If the indicative auction price at the end of the call phase lies outside the applicable thresholds, the auction call phase is extended.

[10] There are only 196 extended volatility interruptions in our sample of 7,899 volatility interruptions.

[11] Our observation period starts with April 2019 as earlier data is not available on the Deutsche Börse A7 market data platform.

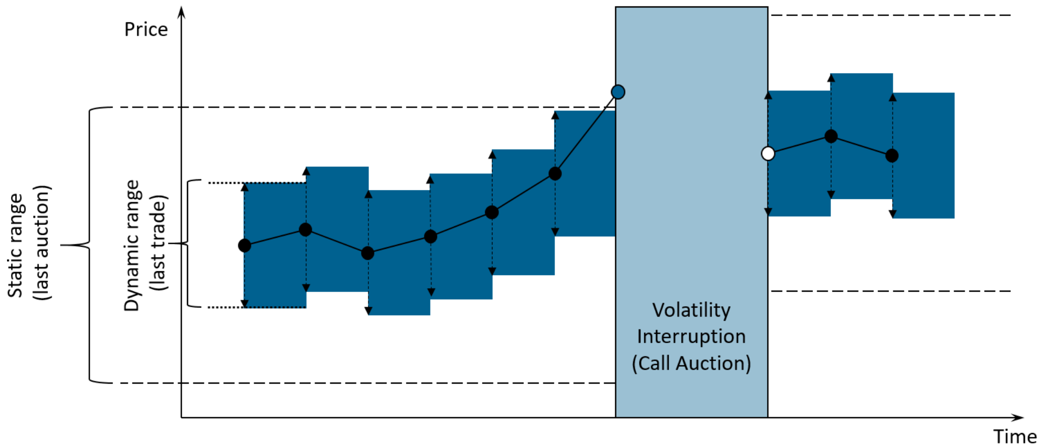

Figure 1: Volatility interruption mechanism on Xetra
The static price corridor is defined symmetrically around the most recent auction price, while the dynamic price corridor is centered around the last trade price. If the next potential execution price lies outside either of these ranges, a volatility interruption is triggered, causing a switch from continuous trading to a call auction phase. Black dots represent execution prices during continuous trading, while the blue dot marks the potential execution price of the incoming order that triggered the interruption. The white dot represents the auction price determined in the volatility interruption. The dynamic price corridor is updated with each execution, the static corridor updates with each new auction price.

In total, we observe 7,899 triggered interruptions for these 49 stocks (subsequently also referred to as DAX40 stocks for simplicity) in this period. For individual volatility interruptions, we further distinguish between an ex-ante "pre-interruption" phase and an ex-post "post-interruption" phase. The selected observation windows of 2 minutes for each phase are short enough to reflect the market's high-frequency nature, yet sufficiently long to capture gradual, longer-term movements that drive the price to the static triggering threshold. For our main approach, we collect 2 minutes of order book, trade, and message data before and after each interruption from the Deutsche Börse A7 market data platform, creating a 4-minute observation period per event. To ensure a complete 4-minute observation window for each event, we exclude 681 interruptions whose 2-minute pre- or post-interruption period overlaps with scheduled auctions (i.e., opening, intraday, or closing auction), and 331 that overlap with other volatility interruptions in the same stock.[12] These cleaning steps result in the final dataset, which comprises 6,887 volatility interruptions. Details of the quantity of observations removed in each cleaning step are depicted in Table 1.

In our empirical study, we utilize Xetra market data with nanosecond granularity, including limit prices and quantities for the top ten levels of the order book, trade executions with corresponding prices and quantities, and order messages detailing submissions, cancellations, and modifications. Furthermore, we incorporate news data from RavenPack (Dow Jones and Press Release Edition), which provides the

[12] These two filtering steps also simultaneously exclude the 196 instances of extended volatility interruptions that occurred during our observation period.

Table 1: Cleaning process and dataset summary

In this table, we present an overview of the data cleaning process and the resulting dataset used for the analysis of volatility interruptions. The first row reports the total number of observed volatility interruptions before cleaning. The subsequent rows list the number of interruptions excluded due to overlaps with scheduled auctions and other simultaneous volatility interruptions. The final dataset, shown in the fourth row, contains the number of remaining volatility interruptions after excluding the overlapping cases. The last row indicates the final sample size as a percentage of the original dataset.

	Number of volatility interruptions
Total number of volatility interruptions	7,899
Overlapping with scheduled auctions	681
Overlapping with other volatility interruptions	331
Final dataset	6,887
Percentage of the original dataset	87%

number of news related to a stock around a volatility interruption together with the relevance and sentiment scores associated with each news item. This dataset from two different sources gives us a comprehensive view of both market microstructure and external informational influences.

As the subsequently introduced machine learning architectures benefit substantially from a fixed input size, we downsample the data to 600 steps before and 600 steps after the event, resulting in a standardized sequence of 1,200 steps per sample. For our main approach, which uses 2-minute pre- and post-interruption periods, this corresponds to a downsampling frequency of 200 milliseconds, which reflects both the high-frequency nature of trading blue chip stocks and sufficient time to observe pre- and post-interruption price trends.[13] Sampling at 200-millisecond intervals is conducted using time-weighted averages for order book data, i.e., prices and volumes on the first ten levels, to preserve as much informational content as possible. As a result, derived metrics such as the relative quoted spread and order book depth are also calculated as time-weighted averages. Trade data is aggregated by counting the number of trades within each 200-millisecond interval and summing their corresponding volumes. News data is aggregated by counting the number of news items within each interval and by taking the mean of the associated sentiment, relevance, and similarity scores. Volatility is computed as the standard deviation of midpoint returns across all 200-millisecond intervals within the respective pre- or post-interruption window. A detailed description of all variables and their aggregation methods is provided in Table 10 in Appendix A. Before deploying machine learning techniques, specifically clustering and deep learning for prediction, we additionally apply min-max normal-

[13] In the robustness tests reported in Section 6, we evaluate alternative observation windows and sampling frequencies while maintaining a constant input length of 1,200 time steps by proportionally adjusting the sampling rate. Specifically, we rerun the entire prediction model using a sampling frequency of 500 ms for a 5-minute window, 300 ms for a 3-minute window, and 100 ms for a 1-minute window around each volatility interruption.

ization to all features, ensuring that each observed interruption is standardized individually (per sample normalization). For bid and ask prices, normalization is based on the minimum and maximum across all price levels within the order book observed in the 200-millisecond intervals around each interruption. The same approach is applied to the volumes across all levels. All other calculated features are normalized separately. While normalization is applied to the entire 1,200-step sequence for the clustering task, it is restricted to the pre-phase for the prediction task to prevent any information from the post-phase leaking into the input data.

We use the entire dataset of volatility interruptions for the clustering process. For the prediction task, we apply a randomized split, dividing the data into 80% train set, 4% validation set, and 16% test set. This approach helps prevent the model from overfitting to the training data and ensures a more balanced subsample, particularly by distributing events from volatile market periods, such as the COVID-19 pandemic or the Russian invasion of Ukraine, more evenly. In addition, it tends to improve model generalization by exposing the model to a broader range of market conditions during the training process.

2.3. Descriptive statistics

Descriptive statistics on the occurrence of volatility interruptions in our dataset are provided in Table 2. During the observation period, DAX40 stocks on Xetra experienced an average of 4.59 volatility interruptions per trading day. However, there are significant outliers, mainly due to market-wide events that triggered numerous interruptions. For instance, the highest number of volatility interruptions on a single trading day was 183. Thus, volatility interruptions are not rare events, even among the most liquid German stocks, but occur frequently, underscoring the importance of identifying and minimizing unnecessary interruptions of trading and price discovery. In terms of variation across stocks, Delivery Hero SE recorded the highest number of volatility interruptions with 449 events, whereas Linde PLC experienced only 13 interruptions, partly due to a slightly shorter observation period, as noted above. A detailed breakdown of the number of interruptions per stock during our observation period is provided in Figure 11 in Appendix A.

Table 2: Descriptive statistics on the observed volatility interruptions in our sample

In this table, we present descriptive statistics on the volatility interruptions observed in our sample. The first row reports the number of volatility interruptions per trading day, with the respective mean, median, minimum, maximum, and standard deviation shown in columns two to six. The second row contains the corresponding statistics for the number of volatility interruptions per stock. The third row provides information on the duration of volatility interruptions in seconds. All values are computed based on the final cleaned dataset.

	mean	median	min	max	std
Volatility interruptions per day	4.59	1.00	0.00	183.00	8.27
Volatility interruptions per stock	140.55	104.00	19.00	449.00	103.36
Duration [sec]	135.02	135.06	120.00	150.00	8.62

On average, volatility interruptions on Xetra lasted 135.02 seconds, closely aligning with the expected duration of 135 seconds, which includes a 120-second auction call phase followed by a random end of up to 30 seconds.

The histogram of volatility interruptions, as illustrated in Figure 2, reveals that these interruptions are occurring regularly on Xetra. It shows significant concentration of volatility interruptions during two global events between 2019 and 2024: the outbreak of the COVID-19 pandemic in March 2020 and the Russian invasion of Ukraine in late February 2022. Both events caused substantial disruptions and economic uncertainty across different industry sectors, resulting in a surge in volatility interruptions during those periods. The frequency of interruptions tends to correlate with periods of broader market stress, as reflected by corresponding spikes in both the number of interruptions and key measures of market quality, such as elevated volatility levels and widening bid-ask spreads.[14]

Figure 3 illustrates the evolution of market quality and trading activity around volatility interruptions in our dataset, thereby already differentiating between necessary and unnecessary interruptions as identified in the subsequent Section 3. The figure shows that liquidity, as measured by the average relative spread, remains relatively stable before the start of an interruption. However, once continuous trading resumes following the auction phase of the interruption, the spread increases substantially by an average of 58% for both necessary and unnecessary interruptions compared to its level immediately prior to the interruption. Interestingly, the average relative spread is on a higher level in case of necessary interruptions compared to those identified as unnecessary. Similarly, liquidity in terms of depth, measured by the euro volume quoted at the first level of the order book (level-1 depth), remains relatively stable, with a slight dip just before the interruption, but then declines by an average of 50% after an unnecessary (32% after a necessary) interruption. Again, liquidity in terms of order book depth is substantially higher before an unnecessary interruption occurs. During the 2-minute observation window following the interruption, liquidity in terms of spreads gradually improves for both necessary and unnecessary interruptions, almost approaching pre-interruption levels. In contrast, market depth shows a partial recovery only after unnecessary interruptions, though it does not return to pre-interruption levels, while it remains persistently lower following necessary interruptions throughout the 2-minute post-interruption period. Consequently, volatility interruptions not only delay price discovery but are also associated with reduced liquidity, leading to higher trading costs for market participants. Also from this perspective, minimizing unnecessary interruptions would be beneficial for market quality.

As expected, volatility increases just before a volatility interruption is triggered. After the interruption, it spikes sharply by 265% after an unnecessary interruption

[14]The reduction in market liquidity during periods of heightened volatility is well documented. Nagel (2012) shows that liquidity providers demand higher expected returns during times of market stress, leading to a withdrawal of liquidity. We do not interpret elevated volatility and illiquidity as competing explanations for periods with a large number of volatility interruptions; rather, they are interrelated dimensions of broader market stress.

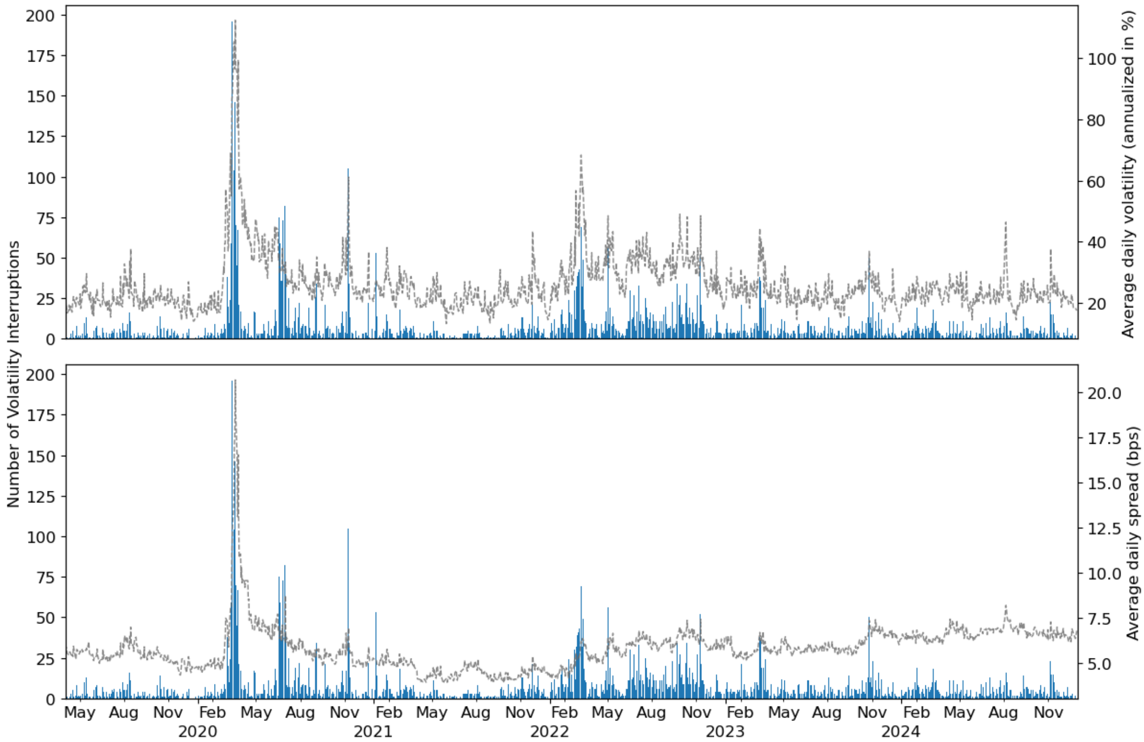

Figure 2: Histogram of volatility interruptions and the evolution of average daily volatility and liquidity in DAX40 stocks
The blue histogram displays the total number of volatility interruptions in DAX40 constituents on Xetra for each trading day during the observation period from 2019 to 2024. The upper panel shows the average daily volatility of DAX40 stocks (dashed line), calculated as the standard deviation of log returns of the mid-price from the lit order book during continuous trading, sampled at a 1-minute frequency. Volatility is annualized using the square root of time rule, assuming 8.5 trading hours per day and 252 trading days per year. The lower panel presents the time-weighted average relative bid-ask spread across all analyzed DAX40 stocks, serving as a proxy for market liquidity.

(and even 307% after a necessary one), in the first few 200-millisecond intervals before gradually settling back to pre-interruption levels over the course of 2 minutes. These descriptive statistics align with the findings of the majority of empirical studies (e.g., Hautsch and Horvath, 2019), which also report increased volatility and reduced liquidity following trading interruptions triggered by circuit breakers.

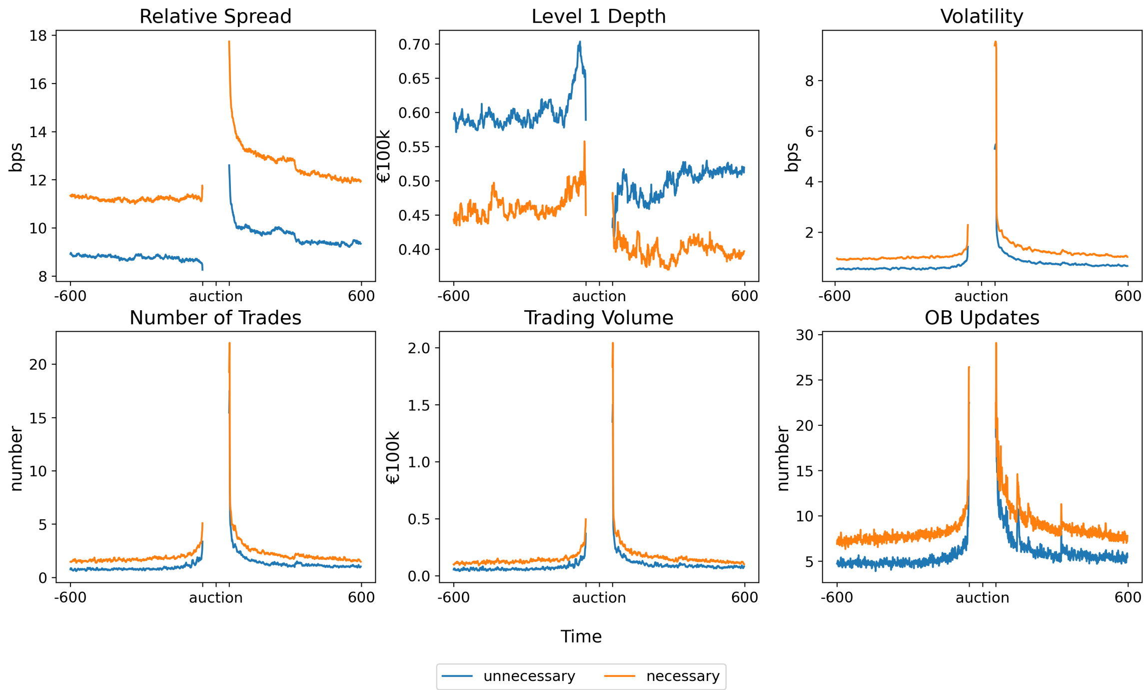

Figure 3: Average liquidity, trading activity, order book activity, and volatility around volatility interruptions

This figure displays the average behavior of key market quality and activity metrics across 600 consecutive 200-millisecond intervals covering the 2-minute continuous trading periods preceding and following a volatility interruption. Specifically, it plots the relative bid-ask spread, depth at the first level of the order book, midpoint return volatility, number of trades, trading volume, and number of order book updates. Trading activity during the call auction phase of the interruption is excluded from the analysis. Since the Xetra order book is not visible during the auction phase triggered by the volatility interruption, the other metrics cannot be calculated for the auction period.

Trading activity, measured by the number of trades and trading volume, experiences spikes both immediately before a volatility interruption is triggered and especially immediately after continuous trading resumes after the interruption. Apart from these two spikes, trading is rather stable around volatility interruptions, with a slight increase in activity in the seconds following an interruption. A similar pattern is observed in order book activity. The spike in trading activity just before a volatility interruption may be driven by market participants responding to new information or reassessing the stock's expected value. It could also result from traders anticipating the interruption and rushing to execute their trades before it takes effect, consistent with the "magnet effect" hypothesis (Subrahmanyam, 1994; Chen et al., 2024). In contrast, the heightened trading activity following the interruption likely stems from participants seeking to execute their intended trades, manage their inventories, and update their orders and quotes—actions they were either unable or unwilling to perform during the auction phase of the volatility interruption.

2.4. Alternative venues

To enhance the comprehensiveness of our analysis and account for potential cross-listing effects, we incorporate data from alternative trading venues for DAX40 stocks. Unlike Xetra, alternative venues in Europe do not implement rule-based circuit breakers comparable to those used on the primary exchange. Instead, they typically rely on discretionary trading halts or enforce price limits that reject executable orders beyond a certain threshold. Moreover, circuit breaker mechanisms in Europe are not coordinated across venues, meaning that continuous trading typically proceeds uninterrupted on these platforms even when a volatility interruption is triggered for the same stock on Xetra.

We obtain trade and order book data from the three largest alternative lit venues offering continuous trading in DAX40 stocks from BMLL Data Lab: Aquis Europe, Cboe Europe Equities, and Turquoise Europe. During the observation period, each operator relocated its European equities trading venue from the UK to the EU due to Brexit. To ensure continuous data coverage, we pair each UK trading book with its corresponding EU successor venue.[15]

Figure 4 presents statistics on trading activity and market liquidity from 15 minutes before to 15 minutes after a volatility interruption on Xetra, across Xetra and the three alternative venues. Across all platforms, trading activity intensifies in the minutes immediately preceding the interruption and peaks in the first minute following it, before gradually returning to pre-interruption levels. This pattern is most pronounced on Xetra and Cboe Europe, the largest alternative venue.

Notably, and in contrast to theoretical predictions suggesting volume migration to unaffected venues during interruptions (Subrahmanyam, 1994), we observe no such migration. On the contrary, trading activity on alternative venues nearly vanishes during volatility interruptions on Xetra. This suggests that market participants largely withdraw rather than reroute trading activity when the main market is paused and the price signal of the reference market is not available.

The mean relative spreads on alternative venues also increase dramatically, by more than fivefold on Cboe Europe compared to pre-interruption levels. This implies that liquidity providers widen their quotes substantially when the main market is paused, and that overall market participation declines. These findings are consistent with prior research on circuit breakers in fragmented European markets (Clapham et al., 2017b; Gomber et al., 2013), suggesting that despite the absence of formal coordination, there is implicit alignment across venues driven by trader behavior and risk aversion.

[15] The cut-off dates are based on the first day of observable trading activity on the new EU venues: 12.11.2020 for Aquis (from AQXE to AQEU), 15.10.2019 for Cboe (from CHIX to CEUX), and 04.01.2021 for Turquoise (from TRQX to TQEX).

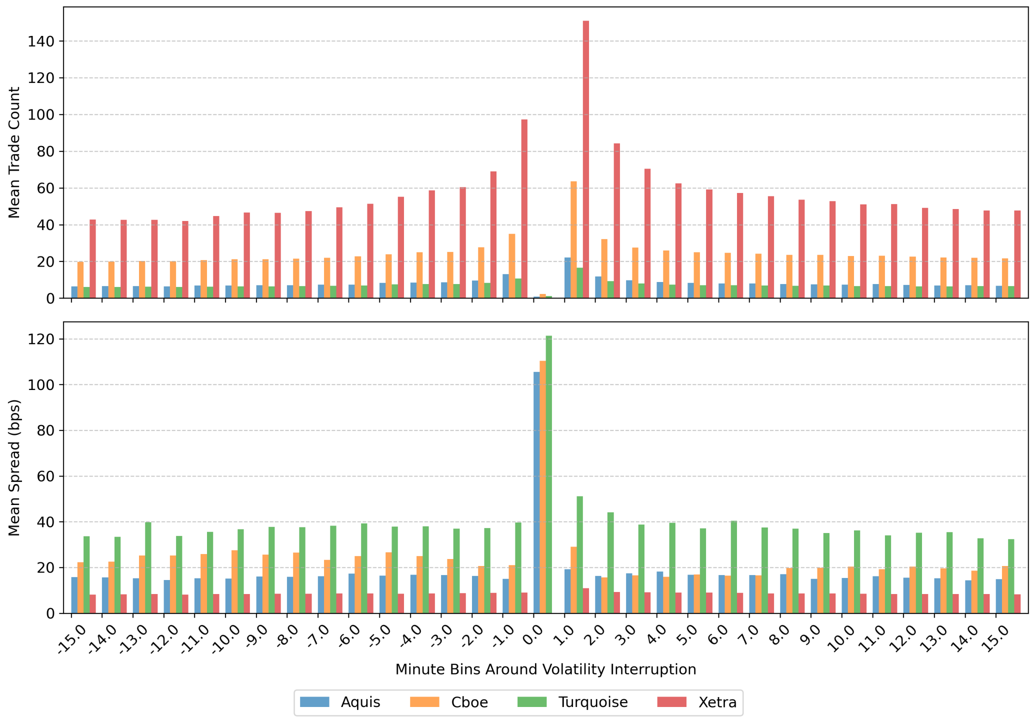

Figure 4: Trading activity and liquidity around volatility interruptions on the alternative venues relative to Xetra
The figure displays the mean trade count (top) and mean relative quoted spread (bottom) per minute interval from 15 minutes before to 15 minutes after a volatility interruption on Xetra, shown across Xetra and the three largest alternative lit venues for DAX40 stocks: Cboe Europe, Aquis Europe, and Turquoise Europe. Interval 0 corresponds to the volatility interruption on Xetra and spans an average duration of 2 minutes and 15 seconds. To ensure comparability across all intervals, trade counts are normalized and reported on a per-minute basis, adjusting for the actual duration of each interval.

3. Clustering and labeling of volatility interruptions

3.1. Methodology

To systematically identify unnecessary volatility interruptions, we analyze historical volatility interruptions by clustering them based on comprehensive LOB data in the 2-minute windows preceding and following each interruption. By using the full set of information contained within the LOB, i.e., limits and volumes of the first ten levels at the bid- and ask-side of the order book, we ensure that only those volatility interruptions characterized by similar market dynamics both before and after the event are grouped together.[16] This approach avoids reliance on a limited set of hand-crafted features from the LOB (e.g., spread or depth measures), offering a more robust categorization of different types of volatility interruptions.

Based on this differentiation of various types of historically observed volatility interruptions, we are able to investigate whether certain clusters exhibit an unintended behavior regarding volatility interruptions, such as a persistent price trend in both the pre- and post-interruption phase or a negative impact on market quality. The presence of persistent price trends would suggest that these interruptions may have merely delayed the natural price formation process, rather than serving as a corrective function for unreasonable price jumps.

To implement this categorization, we utilize an unsupervised clustering algorithm. We base the clustering method on Gaussian Mixture Models (GMMs), which offer a flexible probabilistic approach to clustering. Unlike other clustering algorithms such as k-means, GMMs account for cluster covariance structures and allow for soft assignments, making them well-suited to capture the complex and overlapping nature of market behaviors. Furthermore, the probability estimates provided by the GMM allow us to account for the quality of fit of each observation to its assigned cluster. Specifically, we only include those observations in a cluster which have a probability of more than 90% belonging to this cluster.[17] Given the high dimensionality and temporal dependencies inherent in LOB time series, we start with an autoencoder model to reduce the data to a lower-dimensional set of informative features. This dimensionality reduction ensures that the input to the GMM clustering algorithm consists of time-independent, meaningful features that effectively capture the full range of market dynamics surrounding each interruption.

We base our autoencoder architecture on the DeepLOB model proposed by Zhang et al. (2019), which has demonstrated efficacy in automatically extracting relevant features from LOB data. The encoder component of our model replicates the DeepLOB model, while the decoder component is an exact inversion of the encoder. A schematic illustration of the whole clustering process is given by Figure 13 in Appendix B and

[16] We tested the inclusion of both the number of news items and their sentiment in addition to order book information in the clustering process. However, this extension did not improve clustering results. Due to the relatively low frequency of news events in the 2-minute window surrounding an interruption (news are present in only 7.6% of all cases), the addition of news data introduced noise rather than useful differentiation.

[17] This reduces the total number of observations allocated to the clusters from 6,887 as shown in Table 1 to 5,371 as shown in Table 3.

a more detailed visualization of the architecture for both the encoder and decoder is provided in Figure 15 and Figure 16, respectively. We constrain the latent space between the encoder and decoder to a size of 64[18], ensuring the model extracts only meaningful information from the LOB data. After training the model, we further utilize only the encoder to transform the data for the clustering.

The effectiveness of our clustering results is dependent on the choice of k, the number of clusters. To identify the optimal k, we employ the elbow method, a widely-used technique that sets the value of k where the marginal improvement from adding an additional cluster is largely reduced (Thorndike, 1953; Syakur et al., 2018). This is typically identified as a "bend" or "elbow" in the plot of the sum of squared distances between each sample and its corresponding cluster center for different k.

While circuit breakers are designed to stabilize markets in case of unsubstantiated price jumps, they can adversely impact overall market quality, particularly when triggered by legitimate price movements driven by fundamental information. This is especially evident when volatility interruptions delay the price formation process, such as when a clear and persistent price trend continues in the same direction before and after the interruption. In such cases, the price changes preceding the interruption are not the result of irrational behavior, liquidity shocks, or erroneous orders, but instead reflect the incorporation of new information and evolving market expectations. Hence, the volatility interruption unnecessarily defers price discovery and disrupts trading activity, thereby impairing market efficiency and liquidity. Conversely, a volatility interruption is considered effective if it leads to improved market quality or stabilized, i.e., less extreme, price movements post-interruption.

Accordingly, we classify clusters in which volatility interruptions clearly delay price discovery and worsen market quality as unnecessary. In line with previous research emphasizing the need to evaluate circuit breakers across multiple dimensions (Abad and Pascual, 2013; Hautsch and Horvath, 2019), our methodology accounts for various indicators of market quality to identify instances where interruptions failed to fulfill their intended protective role, with adverse effects outweighing its benefits. Compared to today's circuit breaker mechanisms, which rely solely on rule-based assessments of execution prices relative to predefined thresholds, our approach incorporates multiple relevant dimensions covering both price dynamics and market quality to assess whether an interruption of continuous trading was reasonable or not.

To assess whether observations within a cluster exhibit a persistent price trend, we employ a two-step statistical test setup on the midpoint returns. In the first step, we independently test whether returns before and after the volatility interruption are significantly greater than zero using a one-sided t-test. We evaluate significance using the maximum p-value of the two tests, ensuring that both pre- and post-interruption phases show statistically significant positive returns. This approach guarantees that a persistent upward trend is only identified when both segments individually support

[18]We opted for 64 as we evaluated various other parameter choices (i.e., 512, 256, 128, 64, 32), and 64 was the smallest size with reasonable results as sizes < 64 would result in a much higher loss.

the hypothesis. To test for a persistent downward trend, we apply the same procedure using one-sided tests for negative returns.

In the second step, we compare the results from the upward and downward trend tests by taking the minimum of their respective maximum p-values. This allows us to determine whether the observation is characterized by a statistically significant persistent trend, either upward or downward, surrounding the interruption.

Furthermore, we assess potential negative effects on market quality by conducting a series of t-tests on key metrics. Specifically, we test whether the bid-ask spread is significantly wider in the post-interruption compared to the pre-interruption phase, whether the depth at the top of the order book is significantly lower after the interruption, and whether the volatility of midpoint returns increases following the interruption. These tests complement the preceding analysis of persistent price trends.

A cluster is classified as representing unnecessary volatility interruptions only if all four conditions are met at the 5% significance level: (1) the presence of a persistent price trend, (2) a wider bid-ask spread, (3) reduced order book depth, and (4) increased return volatility post-interruption.

In addition to the statistical framework used to identify unnecessarily triggered volatility interruptions based on persistent price trends and deteriorated market quality, we apply a probabilistic filter to ensure that only interruptions with a high likelihood of being genuinely detrimental to market efficiency are labeled as unnecessary. Specifically, we classify an observation as unnecessary only if its probability of belonging to one of the clusters identified as unnecessary exceeds 90%.[19] This is determined using the posterior probabilities derived from the GMM: for each observation, we sum the probabilities of membership across all clusters deemed unnecessary. An observation is labeled as unnecessary only if this cumulative probability exceeds the 90% threshold.

Further, we acknowledge that while a persistent upward or downward price trend is a clear indication that a volatility interruption has deferred the price formation process, a moderation of the pre-interruption trend in the post-interruption phase may, in fact, align with the objectives of market operators or regulators, who may regard such circuit breakers as effective. However, the statistical test described earlier only detects significant positive or negative midpoint returns and does not account for changes in the steepness of the price trend.

To address this limitation and enhance the robustness of our approach, we introduce an additional measure to assess the dampening effect of volatility interruptions on price trends. Specifically, we estimate the following regression model for each cluster independently:

$$p_c = \beta_0 + \beta_1 t + \beta_2 d + \beta_3 (t \cdot d) \tag{1}$$

[19]We provide the distribution of those probabilities for each cluster in our main clustering approach in Figure 19 in Appendix C. Of the volatility interruptions identified as unnecessary, 2,536 have a probability greater than 90% of belonging to a single cluster classified as unnecessary, while an additional 251 are classified based on a cumulative probability exceeding 90% across multiple clusters identified as unnecessary.

where p_c represents the average midpoint price trend in Cluster c, t denotes the time index ($t \in \{1, 2, ..., 1200\}$), and d is a binary indicator equal to 1 during the post-interruption phase and 0 otherwise. The coefficient of the interaction term β_3 captures the change in the price trend due to the interruption. To quantify the dampening effect, we compute the ratio β_3/β_1, which expresses the reduction in trend steepness as a percentage. For example, a dampening effect of -0.5 indicates that the post-interruption trend is 50% less steep than the pre-interruption trend.

In our analysis, we exclude clusters from being labeled as unnecessary if they exhibit a dampening effect of 50% or more. This conservative criterion ensures that only those interruptions which are highly likely to belong to an unnecessary cluster and fail to meaningfully moderate the price trend are classified as unnecessary. The dampening threshold, here set to 50%, is a flexible parameter that can be adjusted at the discretion of market operators or regulators depending on their policy objectives. We discuss this and further implementation considerations in Section 7.

3.2. Results

To examine which volatility interruptions are classified as unnecessary, Table 3 presents the results of the previously described statistical tests used to evaluate the effectiveness of volatility interruptions across the identified dimensions. These tests are applied to the clustering results obtained from the GMM. We opted for a parameterization of $k = 12$ for the GMM, which consequently results in twelve clusters (we discuss this choice later in this section). The table is organized by cluster classification, with the first five rows corresponding to clusters identified as unnecessary.

It is evident that every cluster of historically triggered volatility interruptions is associated with a statistically significant deterioration in multiple dimensions of market quality, namely, a widening of bid-ask spreads, a reduction in liquidity (measured by order book depth), and an increase in midpoint return volatility during the post-interruption phase. Given this result, the only dimension that clearly distinguishes unnecessary volatility interruptions from others is the persistent continuation of a price trend before and after the interruption.

As shown in Table 3, seven out of the twelve clusters exhibit such a persistent price trend: Clusters 1, 4, 6, 7, 9, 10, and 12. However, two of these clusters (Clusters 4 and 10) show a substantial dampening of the price trend following the interruption, with reductions of 74% and 63%, respectively. As these declines exceed our threshold of a 50% dampening effect[20], we do not classify these clusters as unnecessary. In these cases, the circuit breaker appears to have effectively slowed further price acceleration.

The identification of unnecessary interruptions is thus limited to the remaining five clusters, Clusters 1, 6, 7, 9, and 12, which are listed in the first rows of Table 3. These five clusters are characterized by significant deterioration in market quality and a persistent price trend unaffected by the interruption. Overall, this classification results in 37% of the interruptions in our sample being labeled as unnecessary.

[20] We further discuss the choice of this parameter in Section 7, as this threshold leaves flexibility for market operators or regulators to define what constitutes an effective volatility interruption based on their specific policy objectives.

Table 3: Evaluation of clustering results

This table presents the clustering results from the Gaussian Mixture Model (GMM), along with the outcomes of the statistical evaluation of persistent price trends and declines in market quality. The first column reports the assigned cluster label. The second column shows the p-value from the test for a persistent price trend, while the third column indicates the magnitude of the dampening effect, expressed as a proportion based on β_3/β_1 from Equation 1 (e.g., -0.5 indicates a dampening effect of 50%). Columns four to six report the p-values from the market quality tests. The second-to-last column indicates the number of observations assigned to each cluster after excluding those with a probability less than 90% of belonging into the cluster. Thereby, the number of clustered observations is reduced from 6,887 to 5,371. The final column shows whether the interruptions within each cluster are classified as unnecessary, which also determines the sorting order of the table. One star (*), two stars (**), and three stars (***) indicate a rejection of the null hypothesis at the 90%, 95%, and 99% confidence levels, respectively.

Cluster	Midpoint	Dampening effect	Spread	Depth	Volatility	N	Unnecessary
1	0.00***	−0.45	0.00***	0.00***	0.00***	715	Yes
6	0.01***	0.29	0.00***	0.00***	0.00***	109	Yes
7	0.00***	0.23	0.00***	0.00***	0.00***	606	Yes
9	0.00***	2.07	0.00***	0.00***	0.00***	505	Yes
12	0.00***	−0.39	0.00***	0.00***	0.00***	601	Yes
2	0.48	−1.25	0.00***	0.00***	0.00***	156	No
3	1.00	−1.87	0.00***	0.00***	0.00***	610	No
4	0.00***	−0.74	0.00***	0.00***	0.00***	562	No
5	0.37	−41.62	0.00***	0.00***	0.00***	344	No
8	1.00	−2.68	0.00***	0.00***	0.00***	398	No
10	0.00***	−0.63	0.00***	0.00***	0.00***	445	No
11	1.00	−2.74	0.00***	0.00***	0.00***	320	No

A visualization of the clustering results from the GMM is provided in Figure 5 and Figure 7. These plots illustrate the average midpoint trajectory surrounding each interruption within the identified clusters, both before and after the event, as well as the average bid-ask spread, level-one depth, and midpoint return volatility. While the visualizations depict only four selected measures, representing price and market quality dynamics over a 2-minute window around each interruption, the underlying clustering is based on a much richer representation of the data. Specifically, our clustering approach relies on embeddings derived from the full LOB across the first ten levels, treated as time series data encompassing the entire 2-minute window before and after each interruption. The interruption event is aligned at $t = 0$, with $t = -600$ and $t = 600$ corresponding to two minutes before and after the interruption, respectively, based on a 200-millisecond sampling interval. Volatility is visualized as a marker, representing average midpoint return volatility in both the pre- and post-interruption phases.

Figure 5 presents the five clusters identified as containing unnecessary volatility interruptions. Each of these clusters displays a distinct pattern in which the price trend evident prior to the interruption persists afterward, suggesting that the interruption merely delays an ongoing price formation process. In such cases, the auction phase of the interruption appears ineffective in facilitating efficient price discovery and instead disrupts an established trend.

Clusters 1 and 12 are characterized by pronounced downward and upward midpoint trajectories, respectively, both before and after the interruption. Clusters 7 and 9 show more moderate, yet clearly persistent, midpoint trends in either direction. Despite the variation in price movement intensity, all four of these clusters exhibit similar deteriorations in market quality, namely, wider spreads, reduced depth, and increased volatility following the interruption.

Cluster 6, by contrast, represents a more unique case of unnecessary interruptions. Here, the midpoint remains relatively stable until shortly before the interruption, when a sharp upward movement occurs. This trend continues in the post-interruption phase with only slight attenuation. Notably, the bid-ask spread in this cluster widens significantly more after the interruption than in the other clusters labeled as unnecessary. This cluster, however, only accounts for a smaller proportion of unnecessary volatility interruptions as it only includes 109 observations compared to several hundred in the other unnecessary clusters.

Overall, in each of these five clusters, volatility interruptions do not prevent market inefficiencies resulting from irrational price jumps, instead, they incur costs of delayed price discovery, unrealized trading opportunities, and deteriorated market quality, indicating that the interruptions in these cases do more harm than good.

To illustrate volatility interruptions identified as unnecessary, Figure 6 depicts the midpoint price development for exemplary stocks where an unnecessary interruption

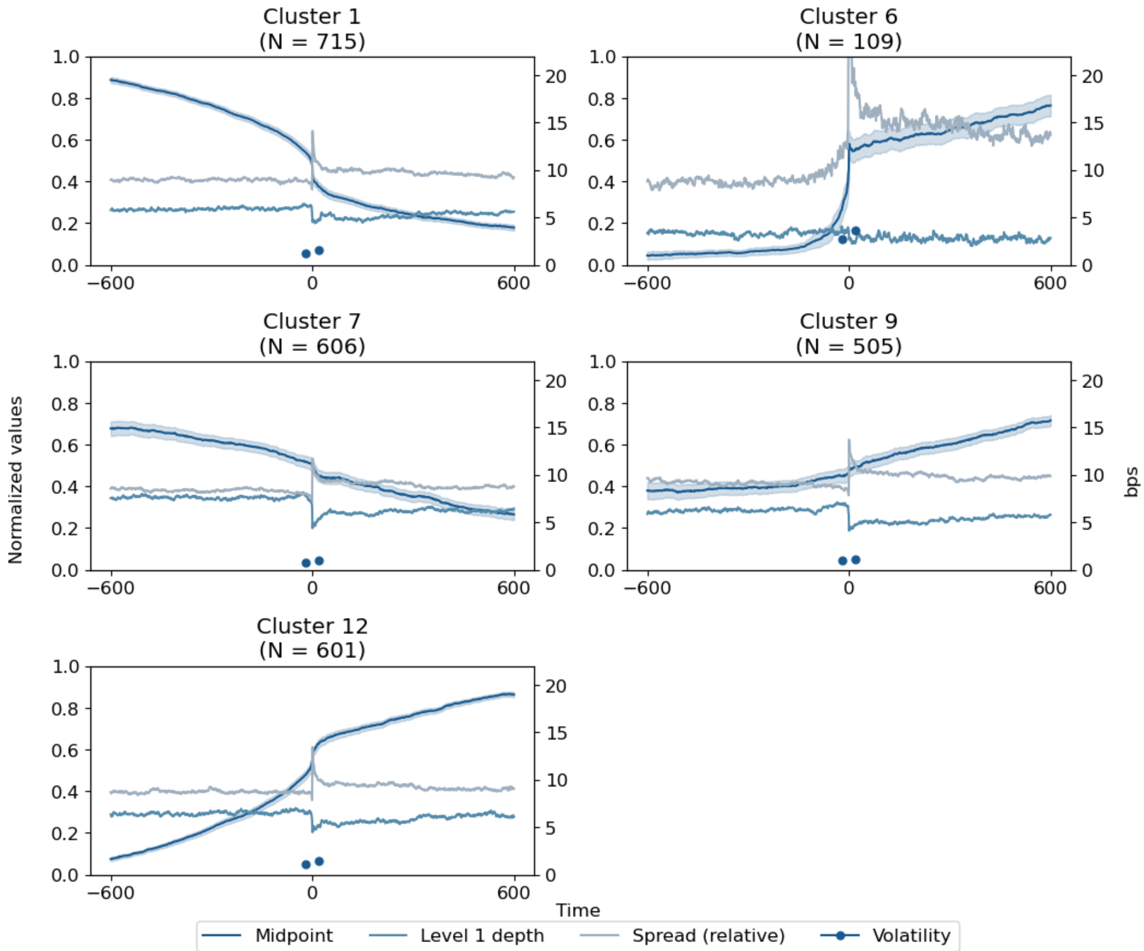

Figure 5: Clustering results for the five clusters identified as unnecessary interruptions
The plot displays the average midpoint trend, level 1 depth, quoted spread, and volatility for all volatility interruptions grouped within each cluster. The auction phase at $t = 0$ is omitted, as no order book information is available and no price determination happens during this period. The number of observations per cluster is shown in parentheses below each subplot title. Midpoint prices and level 1 depth are min-max scaled and plotted on the left y-axis, ranging from 0 to 1. Quoted spread and midpoint return volatility are shown on the right y-axis in basis points. Shaded areas around the average midpoint trend represent 95% confidence intervals.

occurred on a specific day.[21] In each example, the interruption disrupts a persistent price trend that began shortly after the release of significant ad-hoc news related to the respective stock. Consequently, the price changes delayed by the interruption were likely driven by this new information and corresponding shifts in value expectations. For the stocks shown in Panels A and B, the news was positive, resulting in a sustained price increase (a significant upward revision in long-term financial targets for Infineon Technologies and better-than-expected quarterly results for Rheinmetall). In contrast, Panel C features negative news for Deutsche Bank (weaker performance in the credit and derivatives business), leading to a sustained price decline.

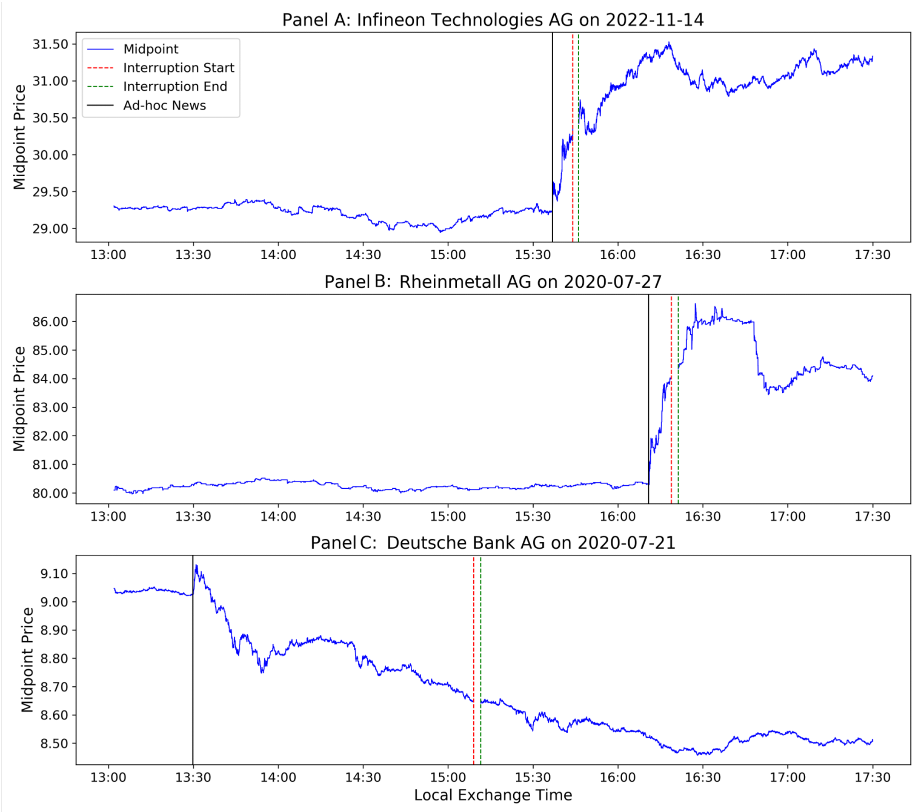

Figure 6: Examples of volatility interruptions identified as unnecessary
Panels A, B, and C show the midpoint price development (y-axis) in blue for Infineon Technologies, Rheinmetall, and Deutsche Bank, respectively, on days experiencing unnecessary volatility interruptions. The start of the volatility interruption is marked in red, while its conclusion and subsequent return to continuous trading is indicated in green. The black vertical line denotes significant ad-hoc news releases. The x-axis displays time in local exchange time.

All remaining clusters, shown in Figure 7, are not classified as unnecessary. Based on our statistical evaluation, these clusters do not exhibit signs of delayed price forma-

[21]Since all volatility interruptions in this exemplary sample occurred in the afternoon, we only plot the period from the resumption of continuous trading after the intraday auction to market close. No additional volatility interruptions or ad-hoc disclosures occurred for these stocks on the respective dates. Ad-hoc news reports are included in the news data from RavenPack used in this study.

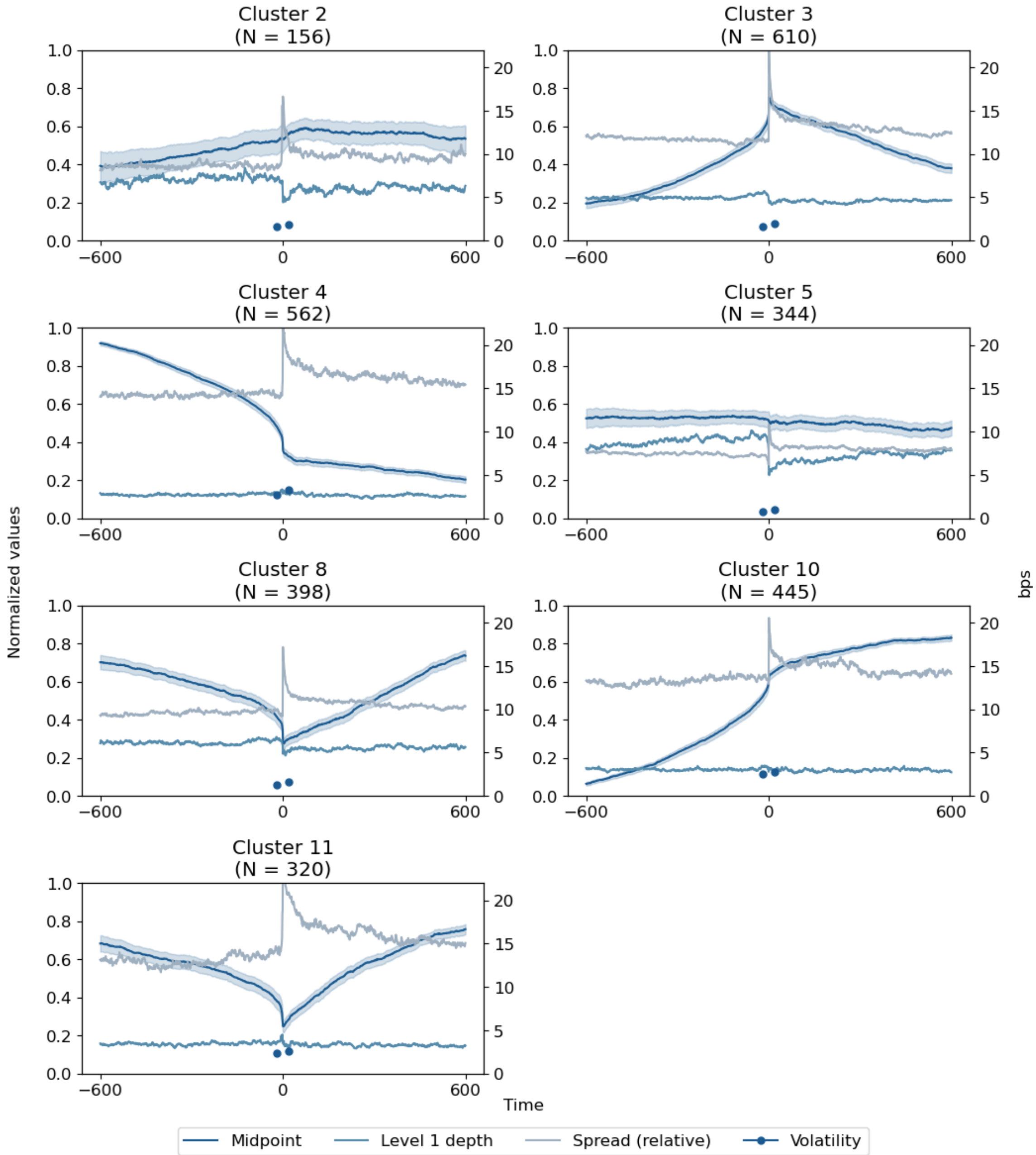

Figure 7: Clustering results for the remaining clusters not identified as unnecessary interruptions The plot displays the average midpoint trend, level-1 depth, quoted spread, and volatility for all volatility interruptions grouped within each cluster. The auction phase at $t = 0$ is omitted, as no order book information is available and no price determination happens during this period. The number of observations per cluster is shown in parentheses below each subplot title. Midpoint prices and level-1 depth are min-max scaled and plotted on the left y-axis, ranging from 0 to 1. Quoted spread and midpoint return volatility are shown on the right y-axis in basis points. Shaded areas around the average midpoint trend represent 95% confidence intervals.

tion, as indicated by the absence of a continued price trend following the interruption, a key characteristic of the earlier clusters identified as unnecessary. While volatility interruptions in the remaining clusters are also associated with lower post-interruption liquidity and elevated volatility (see Table 3), they appear to contribute positively to price stability. Specifically, they are followed by either a stabilization or reversal of the pre-interruption price movement, or at least a substantial dampening of the ongoing price trend compared to pre-interruption dynamics. Consequently, interruptions in these clusters seem effective in "cooling down" the market during periods of extreme price movements, giving participants time to reassess available information and determine the validity of price movements, as described by Ma et al. (1989).

This behavior is particularly evident in Clusters 3, 8, and 11, where the extreme price trends observed before the interruption are almost entirely reversed in the post-interruption phase. This suggests that, during the pre-interruption phase, the market may have either been uncertain about the true value of an asset, overreacted to certain news, or experienced substantial market impact from the execution of a large order. In these cases, the interruption phase allowed for a reassessment, showing that the interruption served as a valuable safeguard and contributed directly to price discovery.

Clusters 4 and 10 display a similar behavior, though without the trend-reversing effect observed in the other clusters. Instead, these clusters exhibit a trend-breaking effect, where the price trend stabilizes around the auction price level in the post-interruption phase or at least is dampened to a reasonable degree. This suggests that the volatility interruption effectively contributed to the price formation process, as prices tend to remain close to the price level of the volatility interruption.

The remaining two clusters, 2 and 5, are characterized by relatively steady price trajectories without a clear upward or downward trend. Despite the absence of sustained price movements, interruptions in these clusters were likely triggered by sudden, extreme price jumps occurring within otherwise calm market conditions. In such cases, the volatility interruption appears to have served a valuable corrective function by safeguarding the price discovery process from potentially erroneous or irrational market movements leading to a stable price level after the interruption.

The choice of the number of clusters ($k = 12$) merits further discussion. As outlined earlier, we apply the elbow method to determine the point at which adding additional clusters results in only marginal improvements in clustering performance. The results of this procedure are shown in Figure 18 in Appendix C. This figure presents the sum of squared distances between each observation and its assigned cluster center across a range of values for k ($k \in \{2, 3, ..., 30\}$).

To robustly identify the elbow point, we apply two established methods: (1) detecting the point at which the second derivative of the interpolated curve reaches its maximum, and (2) determining the point with the greatest perpendicular distance to the diagonal connecting the first and last entries of the curve. Both approaches are widely used for identifying the inflection point of the curve, and using both provides a more reliable assessment. The results suggest that an optimal balance is achieved at $k = 9$ or $k = 11$, depending on the method.

Given that increasing the number of clusters beyond this point does not incur significant computational costs, aside from added complexity in handling and visualization[22], we interpret the outcome of the elbow method as a lower bound for selecting k. Additionally, we argue that an even number of clusters is preferable in our context, as we expect a degree of symmetry between clusters characterized by upward and downward price movements.

Based on these considerations, we select $k = 12$, which satisfies the lower bound suggested by the elbow method while aligning with our expectation of directional symmetry. This configuration introduces a sufficient number of clusters to capture the heterogeneity in market dynamics around volatility interruptions without introducing excessive redundancy. A sensitivity analysis on the effect of varying k on the identification of unnecessary volatility interruptions is provided in Table 14 in Appendix D showing only a moderate effect of the choice of other k for the final labeling into unnecessary interruptions. With an average of 82%, the vast majority of volatility interruptions identified as unnecessary remain consistent across different cluster specifications when using $k \in \{10, 12, 14, 16, 18\}$.

4. Drivers and implications of unnecessary volatility interruptions

4.1. Methodology

We investigate the market conditions that lead to unnecessary interruptions using probit regression models. These models apply different sets of independent variables to explain the binary dependent variable, which indicates whether an interruption was identified as unnecessary. Given that our dependent variable is binary, probit regression is an appropriate choice. It effectively models binary outcomes by producing a continuous output that represents the probability of the dependent variable belonging to the target class (here: unnecessary volatility interruptions).

Our first model focuses on the *market quality* conditions prior to the interruption, with the goal of determining whether liquidity, volatility, or trading activity related factors influence the likelihood of an interruption being unnecessary. Therefore, we incorporate the average relative spread, order book depth, trading volume, midpoint volatility, and number of order book messages prior to each interruption in this analysis.

Factors beyond market data, such as news events, are often associated with price adjustments. Therefore, we introduce a second probit regression model that incorporates *news and contextual factors* to account for macroeconomic conditions prior to the interruption. Inputs for this model include the number and relevance of recent news articles related to the issuer of the affected stock, the number of necessary and unnecessary volatility interruptions across all DAX40 stocks in the respective market

[22] Technically, clustering quality continues to improve as k increases, up to the theoretical extreme where each observation forms its own cluster.

(Xetra) within the preceding hour[23], the fast market indicator[24], trade data from three alternative venues[25], and the proximity of the last price prior to the interruption to the estimated static triggering threshold.[26] A comprehensive overview of the variables used in both models, including their calculations and a brief explanation of their informational content, is provided in Table 10 in Appendix A.

To address the time-dependent nature of our input data, which covers two minutes before each volatility interruption, we use an exponentially decaying average (or sum for count- and volume-based variables such as the number of order book messages and trading volume). This method assigns progressively less weight to data points further from the interruption, ensuring that more recent information has a greater influence on the model's predictions. By transforming the time series data into one time-independent measure for each interruption, this approach makes it compatible with the probit regression framework.

Following this approach, we set up the first probit regression - the *market quality* model - using the following equation:

$$\begin{aligned}\Pr\left(y_i = 1 | X_i\right) = \Phi\,(&\alpha + \beta_1 \cdot rel\_spread_i + \beta_2 \cdot level\_1\_depth_i \\ &+ \beta_3 \cdot trade\_volume_i + \beta_4 \cdot midpoint\_return\_vola_i \\ &+ \beta_5 \cdot message\_count_i)\end{aligned} \tag{2}$$

where $i \in \{1, 2, ..., 6887\}$ is the index for each observation, y is the binary target, Φ is the cumulative normal distribution and $X_i = \{rel\_spread_i,\ level\_1\_depth_i,\ trade\_volume_i,\ midpoint\_return\_vola_i,\ message\_count_i\}$ is the list of independent variables.

[23]The inclusion of the number of volatility interruptions separately counted for those identified as necessary and unnecessary across all DAX40 stocks within the hour preceding an interruption should capture the influence of broader market-wide events, such as the COVID-19 pandemic, which can trigger simultaneous price adjustments across multiple stocks.

[24]The Xetra market supervision department can define a so called fast market for all stocks for a specific day if unusually high volatility is anticipated. This could be the case if major macroeconomic events or crucial announcements are expected or occurred before market opening, which affect the entire market. Then, the price corridors for the triggering of volatility interruptions are doubled.

[25]As described in Section 2, we use trade data from Aquis Europe, Cboe Europe Equities, and Turquoise Europe.

[26]Since the actual price thresholds are not publicly disclosed, we reverse-engineer the static threshold for each stock by observing the maximum price deviation from the last auction price that did not trigger an interruption. Given that the market operator may adjust these thresholds in response to overall market volatility, especially during volatile periods, we independently estimate the static price range for each stock and month to account for potential adjustments. Figure 10 in Appendix A shows the resulting approximated static price ranges. The distance of the last price prior to the interruption to the static threshold offers insights into whether a significant price jump triggered the volatility interruption or if only minor price changes were sufficient to initiate it. This distinction can provide valuable information for differentiating between unnecessary and relevant interruptions.

The second regression - the *news and contextual factors* model - is set up by the following equation:

$$\begin{aligned}
\Pr(y_i = 1|X_i) = \Phi\,(&\alpha + \beta_1 \cdot news_count_i + \beta_2 \cdot news_relevance_i \\
&+\beta_3 \cdot news_count_i \times news_relevance_i \\
&+\beta_4 \cdot unnecessary_interruptions_i \\
&+\beta_5 \cdot necessary_interruptions_i \\
&+\beta_6 \cdot fast_market_dummy_i \\
&+\beta_7 \cdot distance_to_static_barrier_i \\
&+\beta_8 \cdot trade_count_alt_venues_i)
\end{aligned} \tag{3}$$

where $X_i = \{news\_count_i,\ news\_relevance_i,\ unnecessary\_interruptions_i,$ $necessary\_interruptions_i,\ fast\_market\_dummy_i,$ $distance\_to\_static\_barrier_i,\ trade\_count\_alt\_venues_i\}$ is the list of all independent variables in the *news and contextual factors* model.

As a third model, we introduce a comprehensive regression model that incorporates all independent variables used in the previously discussed models. This *full* model aims to integrate a broad range of factors, including market quality metrics, news, and other contextual variables, to explain the likelihood of an unnecessary volatility interruption ex-ante. By encompassing all relevant factors, this *full* model is anticipated to offer a more robust and accurate assessment.

While the primary objective of this section is to identify the key drivers that lead to the triggering of unnecessary volatility interruptions, the predictive performance of the proposed empirical models also serves as a baseline for assessing the feasibility of ex-ante prediction of such events. To evaluate each model's capability in accurately predicting the probability of a volatility interruption being unnecessary using only ex-ante information, we apply common classification metrics to the models' outputs such as accuracy, precision, and recall: Accuracy measures the overall agreement between predicted labels and true labels. Precision for a specific class measures the proportion of true positives among all predicted positives, reflecting the model's likelihood of correctly predicting the target class (here: unnecessary volatility interruptions). Recall measures the proportion of true positives identified out of all actual positives in the target class, indicating the model's ability to capture all relevant instances of unnecessary volatility interruptions.

In line with standard model evaluation practices, we conduct an out-of-sample test based on randomly splitting the dataset into training, validation, and test subsets as described in Section 2. For the described regression models, we use both the training and validation sets for training as we do not have to control for model overfitting given the limited number of parameters in the probit models. In addition to standard performance metrics, we report the highest precision score achieved at a minimum recall score of 0.2. This score is obtained by varying the decision threshold applied to the model's probabilistic output (e.g., classifying values above 0.5 as unnecessary) and identifying the highest precision attainable while maintaining at least 0.2 recall. While all other classification metrics are reported using the default threshold of 0.5, the precision-at-recall score offers insights into how alternative thresholds may improve precision for a fixed recall level.

The primary goal of this study is the accurate identification of unnecessary volatility interruptions to minimize these interruptions and mitigate their negative effects. Therefore, an ideal classification model should maximize both precision and recall when classifying unnecessary interruptions. High precision ensures that when the model predicts an interruption as unnecessary, it is highly likely to be correct. High recall ensures that the model identifies as many unnecessary interruptions as possible. However, prediction models regularly face a trade-off between precision and recall. For our task, i.e., improving circuit breaker mechanisms, precision is the key metric as a high precision ensures that the model predicts an unnecessary interruption only when the prediction is highly likely to be correct, thereby minimizing the risk of false positives. In a real-world implementation, it is less critical to capture every unnecessary interruption (recall), as each correctly identified and avoided interruption would already improve market efficiency relative to the status quo. However, false positives could disrupt price continuity and compromise overall market stability. Therefore, the model's precision must be prioritized to reduce the risk of misclassifying necessary interruptions as unnecessary, which is crucial to prevent welfare losses caused by large transitory price swings due to, e.g., short-term liquidity crashes or erroneous trading algorithms. We discuss the welfare implications of preventing unnecessary volatility interruptions under the existence of false positives in Section 5.3.

4.2. Results - drivers of unnecessary interruptions

Table 4 presents the outcomes of the regression models. The coefficients in the table represent marginal effects, indicating the absolute increase (decrease) in the probability of the dependent variable given a variation for each explanatory variable.

The probit regressions help identify the market conditions under which unnecessary volatility interruptions occur. In the first model, trading volume does not show a significant effect, indicating that unnecessary interruptions may arise during both high and low activity periods. This suggests that trading volume is not a primary driver. In contrast, liquidity conditions play a more prominent role: narrower bid-ask spreads and greater depth at the top of the order book are significantly associated with a higher likelihood of an interruption being unnecessary. In such cases, the cause behind the price change triggering the interruption was not a short-term liquidity shock, which is one reason why circuit breakers are in place.

Additionally, higher pre-interruption price volatility decreases the probability of an interruption being unnecessary, supporting the idea that circuit breakers are more justified during turbulent market conditions. Conversely, unnecessary interruptions tend to occur in calmer, more stable phases where markets are functioning efficiently and external intervention is less warranted.

Necessary interruptions are also linked to elevated order book activity, reflecting increased order submissions, modifications, or deletions, which is typical of volatile or uncertain markets. In contrast, unnecessary interruptions are often triggered during more orderly phases, reinforcing the view that they may disrupt rather than support efficient price discovery.

Table 4: Marginal effects of probit regression models

This table presents the marginal effects from the probit models, where the dependent variable is a binary indicator of unnecessary volatility interruptions. Independent variables are listed in the first column. Marginal effects represent the absolute change in predicted probability due to a one-unit change in each variable. All variables are exponentially weighted means over the two-minute window before the interruption, except for trading volume, news count, trade count on alternative venues, and volatility interruption count, which are exponentially summed. The fast market dummy equals one if active. Counts of unnecessary/necessary interruptions reflect all such events in the preceding hour. Relative spread and volatility are in basis points (bps); level-1 depth and trade volume in hundreds of thousands of euros; message count in thousands. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

	Market Quality	News and Contextual Factors	Full Model
rel_spread	−0.0044***		−0.0027***
	(0.0010)		(0.0010)
level_1_depth	0.0722***		0.0699***
	(0.0124)		(0.0123)
trade_volume	−0.0001		−0.0005
	(0.0003)		(0.0003)
midpoint_return_vola	−0.0604***		−0.0559***
	(0.0087)		(0.0089)
message_count	−0.0173***		−0.0167***
	(0.0025)		(0.0026)
news_count		−0.0008**	−0.0007**
		(0.0003)	(0.0003)
news_relevance		−0.1583***	−0.0818**
		(0.0004)	(0.0004)
news_relevance_interaction		0.0709*	0.0792**
		(0.0368)	(0.0356)
unnecessary_interruptions		0.0046***	0.0033***
		(0.0011)	(0.0011)
necessary_interruptions		−0.0024***	−0.0016***
		(0.0004)	(0.0004)
fast_market_dummy		−0.0002***	−0.0001**
		(0.0001)	(0.0001)
distance_to_static_barrier		0.2762***	0.0678*
		(0.0384)	(0.0400)
trade_count_alt_venues		−0.0011***	0.0006**
		(0.0002)	(0.0003)
Pseudo R2	0.0623	0.0333	0.0691
Accuracy	0.6138	0.5730	0.6074
Precision	0.5670	0.4265	0.5464
Recall	0.3210	0.0629	0.3579
Precision at recall 20	0.5670	0.5033	0.5536

In the second regression model, the analysis of news and contextual factors shows that the presence of relevant news, captured by the interaction between news count and relevance score, significantly increases the likelihood of an interruption being unnecessary. This supports the hypothesis that such interruptions can hinder the price discovery process by delaying the incorporation of new information into prices.

The proximity of the current price to the static price limit also significantly increases the probability of an interruption being classified as unnecessary. This indicates that the existing rule-based mechanism, which triggers interruptions solely when the next potential price exceeds predefined bounds, lacks the flexibility to differentiate between abrupt market disruptions and gradual price movements occurring near the threshold.

Another significant factor reducing the probability of unnecessary interruptions is the activation of the fast market indicator. When high volatility is anticipated, this indicator expands price limits, and interruptions under these conditions are more likely to be necessary.

Market context further influences interruption classification: a higher number of unnecessary interruptions in the preceding hour increases the likelihood of another unnecessary interruption, while a higher number of necessary interruptions decreases it. This suggests that clusters of similar types of interruptions tend to occur in distinct market regimes: unnecessary ones during stable periods and necessary ones during volatile phases. When one regime dominates, the likelihood of corresponding interruptions rises.

Regarding cross-listing effects, the trade count on three alternative venues before the interruption yields mixed results: a negative effect in the second regression model and a positive one in the full model. As such, no definitive conclusion can be drawn about the role of cross-market activity in predicting unnecessary interruptions.

The results from the full model, which combines both individual models, largely confirm the earlier findings, underscoring their robustness. The only notable inconsistency remains the ambiguous role of trade activity on alternative venues.

The classification metrics reported at the end of Table 4 indicate that the first two models exhibit limited predictive power, with accuracy scores below 62%. This performance is comparable to a naive baseline that simply predicts the majority class in the imbalanced test set, where 58.2% of the observations are classified as not unnecessary. Similarly, the full model achieves an accuracy of just 60.7%, falling short of a satisfactory threshold for predictive reliability.

When assessing each model's best precision at a minimum recall of 20% (as shown by the "precision at recall 20" metric), none exceeds a precision of 57%, highlighting their limited effectiveness in correctly identifying true positives, i.e., unnecessary interruptions. In fact, precision remains below 57% across all models, and recall is even lower, peaking at 36% for the full model.

In summary, none of the models demonstrate a reliable ability to predict, ex-ante, whether an interruption is unnecessary. These results suggest the need for model refinement or alternative modeling strategies to improve predictive performance.

Overall, unnecessary volatility interruptions are more likely to occur during active price formation processes, especially in the presence of relevant news, when overall liquidity is high and volatility is low. These market conditions suggest that current price changes are supported by market participants, and interruptions in such situations unnecessarily disrupt ongoing price discovery processes. Furthermore, triggering interruptions when prices are close to the static price limit is more likely to be unnecessary, highlighting the inherent problem of simplistic, rule-based circuit breakers.

However, our analysis shows that predicting unnecessary volatility interruptions ex-ante based on our probit regressions is neither feasible nor adequate for improving the circuit breaker mechanism. While the probit regressions offer insights (explainability) into the factors driving different types of volatility interruptions, the complexity of the market conditions in which these interruptions occur goes beyond the capacity of such models to capture dynamic interactions between various factors.

Therefore, we propose a comprehensive deep learning model designed to capture and model the complex relationships between multiple market quality and contextual factors in Section 5.

4.3. Implications of unnecessary interruptions for market quality

Having established that unnecessary volatility interruptions are more likely to be triggered under liquid market conditions, it is important to examine whether their market implications differ from those of necessary interruptions. As shown in Section 2 and illustrated in Figure 3, the triggering of any volatility interruption is, on average, associated with a deterioration in market quality, consistent with the findings of Hautsch and Horvath (2019). However, Figure 3 also indicates that the widening of bid-ask spreads is less pronounced for interruptions classified as unnecessary compared to those deemed necessary. In this section, we aim to formally test whether the impact on market quality differs significantly depending on the type of the interruption. Therefore, we run the following regression model shown in Equation 4:

$$\Delta MarketQuality_i = \alpha + \beta_1 \cdot unnecessary_i + \delta' \cdot \mathbf{X}_i + \epsilon_i \tag{4}$$

Where $\Delta MarketQuality_i$ is the difference in the respective market quality measure (i.e., average relative spread, average depth at the top of the order book, and midpoint return volatility measured as the standard deviation of all 200-millisecond intervals) between the post- and the pre-interruption period (e.g., $\Delta rel\_spread_i = rel\_spread_{i,post} - rel\_spread_{i,pre}$). The term $unnecessary_i$ is a dummy variable that equals 1 if the interruption i is classified as unnecessary, and 0 otherwise. The control vector X_i includes all explanatory variables used in the full probit model (see Table 4), but computed using standard time-weighted averages instead of exponential averages to ensure consistency with the calculation of market quality changes.

The results reported in Table 5 show that unnecessary interruptions are associated with smaller declines in liquidity and less accelerating levels of volatility. The coefficient on $unnecessary_i$ is negative and significant for the change in relative spreads, indicating that spreads widen significantly less after unnecessary interruptions. Specifically, the increase in the relative spread is 0.3 bps lower compared to the average spread increase of 1.27 bps after an interruption, corresponding to a reduction of

26.9%. Similarly, the reduction in order book depth is significantly less pronounced for unnecessary interruptions (by 54.9% relative to the unconditional average reduction of €8,030). Finally, the increase in volatility after an interruption is smaller for unnecessary interruptions, corresponding to a reduction of 17.5% relative to the average increase in volatility after an interruption. Together, these findings suggest that unnecessary volatility interruptions tend to come with less adverse effects on market quality compared to interruptions deemed necessary. This suggests that the market situation after the interruption mirrors the unnecessary nature of the interruption.

Table 5: Impact of unnecessary interruptions on market quality

This table reports the results of the regression specified in Equation 4, which tests whether unnecessary volatility interruptions affect market quality differently from interruptions classified as necessary. The dependent variables include Δrel_spread, the change in the time-weighted average relative spread, $\Delta level\_1\_depth$, the change in the time-weighted average depth at the top of the order book, and $\Delta return\_vola$, the change in volatility measured as the standard deviation of midpoint returns over 200-millisecond intervals. Each variable is computed as the difference between the post- and pre-interruption phases ($Post - Pre$). The row labeled $Average\Delta$ reports the unconditional average of these differences across all volatility interruptions. The key explanatory variable, *unnecessary*, is a binary indicator equal to 1 if the interruption was classified as unnecessary. The control variables mirror those used in the full probit model reported in Table 4, but are calculated using standard time-weighted averages (rather than exponential averages) to ensure consistency with the dependent variable definitions. For brevity, coefficients of the control variables are omitted. The full output is provided in Table 12 in Appendix C. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

	Δrel_spread	$\Delta level\_1\_depth$	$\Delta return\_vola$
$Average\Delta$	1.27bps	−0.08(€100k)	0.32bps
constant	3.6333***	0.2615***	0.6777***
	(0.3536)	(0.0211)	(0.0590)
unnecessary	−0.3268***	0.0434***	−0.0588***
	(0.1168)	(0.0070)	(0.0195)
controls included	Yes	Yes	Yes
R-squared	0.0730	0.5704	0.1430
R-squared Adj.	0.0711	0.5696	0.1413
Observations	6887	6887	6887

5. Prediction of unnecessary volatility interruptions using deep learning

5.1. Methodology

Building on the results presented in Section 4, we propose a deep learning model designed to predict the likelihood of a volatility interruption being unnecessary, using only information available prior to its triggering. Unlike the models employed in the previous section, our deep learning approach can effectively model complex spatial and temporal non-linear relationships among various input factors. This capability eliminates the need to aggregate input time series into a single value per feature, as required in the probit regressions using exponential aggregations. Moreover, it removes the necessity of pre-selecting potentially relevant independent variables, as the model's architecture inherently identifies and extracts useful features.

Consequently, our deep learning model can process raw information directly without relying on pre-calculated measures. This approach enables us to leverage large volumes of unprocessed data preceding each volatility interruption, capturing the full spectrum of dynamics and relationships among various factors that influence whether the interruption would be unnecessary.

The classification model incorporates two distinct data streams: LOB data and time-dependent news and contextual information. We further differentiate the model setup into two prediction setups: (1) the market data model, and (2) the full model, which distinguish themselves solely based on the input data used. The market data model only uses market data for the prediction and is therefore easier to implement in terms of data requirements. The full model uses additional data, especially news and trade data from alternative trading venues. We discuss practical considerations to implement the models in Section 7. The following subsections describe the deep learning architecture, thereby focusing on the full model, as the implementation of the reduced market data model only requires leaving out the non-market data for the model input data.

The LOB data stream

The LOB data stream comprises 600 snapshots (equivalent to two minutes sampled at a 200-millisecond frequency) of LOB data, each containing ten levels of price and quantity information for both the bid and ask sides. The model structure for the LOB data stream is inspired by the architecture of DeepLOB (Zhang et al., 2019). The initial part of the architecture features a CNN comprising multiple neurons responsible for convolution operations. The CNN aims to extract useful features from single LOB snapshots, potentially encompassing convolutions of bid- and ask-side price levels or comparisons of prices and quantities. This distinguishes it from the probit regressions (Section 4), as it autonomously extracts useful features by design from the order book data, removing the need for pre-calculated and time-aggregated measures. Consequently, it should lead to more meaningful features to be extracted and later used, since the model itself employs the features needed for successful classification or dimensionality reduction.

Subsequent to the feature extraction by the CNN layers, the architecture employs an inception module, focusing on time-wise convolution rather than convolution within a single orderbook snapshot. This module is designed to extract features based on the ones previously computed over a specific time frame, summarizing them into one value. For instance, it could calculate the maximum spread over the last five time steps. The extracted features are then fed into an LSTM layer, which captures temporal changes in these features essential for the time-dependent characteristics of the LOB data stream. Two key modifications have been made to the original DeepLOB architecture of Zhang et al. (2019): the dropout layer's activation probability is reduced to 0.02, which has demonstrated improved performance. Additionally, the architecture is truncated after the final LSTM layer to produce a feature vector for concatenation with the news and contextual data stream.

The news and contextual data stream

Analogous to the regression-based *news and contextual factors* model in Section 4, the news and contextual data stream in the deep learning model is designed to capture information that is not directly observable in market data but may influence whether an interruption is deemed unnecessary. This stream covers the same types of information as those used in the second probit regression in Section 4. However, instead of aggregating this information, as it was done for the regressions, we utilize the raw time series data as input for the deep learning model.

Given the deep learning model's higher expressive power and its ability to autonomously extract relevant features for prediction, we include a broader array of information in the news and contextual data stream beyond the pre-selected measures used earlier. This enables the incorporation of a wide range of contextual factors during the model training process, aiming to capture as many potential dynamics and relationships as possible to enhance prediction accuracy.

In addition to the factors used in the regression model, we include the number of past interruptions for the affected stock besides the number of interruptions in all DAX40 stocks, aiming to capture stock-specific high-volatility dynamics. Our hypothesis is that multiple interruptions in the same stock may indicate unnecessary interruptions triggered by a substantial but relevant change in the fundamental value of the stock. We also incorporate the number of interruptions at both the stock and market level (i.e., across all DAX40 stocks) over the past 24 hours to account for long-term dynamics. For all variables capturing the number of past interruptions, we distinguish between those identified as unnecessary and those that are not, which enables the model to differentiate between the types of past interruptions and assess their potentially distinct informational contributions when predicting the necessity of future interruptions.[27] To enrich the trade data, we add the number of executed trades and their average trade volume. In terms of news data, we include the sentiment score of each news item and the similarity of new information to past news, which helps to determine whether the news is recent or if its content may already be reflected in prices. Furthermore, we include order message data, encompassing the number of order submissions, cancellations, and modifications, to provide a detailed view of activities within each order book update. Lastly, we factor in the current minute of the observation, recognizing that key announcements by central banks or similar entities often occur at specific times, such as on the quarter, half, or full hour.[28] Furthermore,

[27]Even in a real-world implementation, where unnecessary interruptions might ideally not be triggered, such events would still be identifiable in the historical data. A market operator's systems can track whether an interruption would have been triggered under existing rules, even if it was ultimately prevented based on our proposed model's classification as unnecessary. Therefore, it remains both feasible and meaningful to condition on the historical occurrence of unnecessary interruptions in the prediction process.

[28]For example, the European Central Bank's monetary policy decisions are published in a press release at 14:15 CET (equals local time for the trading venue Xetra) followed by a press conference at 14:45 CET. See https://www.ecb.europa.eu/press/govcdec/mopo/html/index.en.html.

we augment the data stream by incorporating information on trades executed on three alternative trading venues. This addition enables us to capture trading activity beyond the main market in the lead-up to an interruption as trading behavior on other venues may offer valuable insights into the causes of volatility observed on the primary market. For instance, if all markets exhibit similar dynamics and activity patterns, this suggests that the main market is functioning properly. Conversely, a significant discrepancy between the main market and alternative venues may indicate the presence of erroneous behavior or market dysfunction on one of the platforms.

In summary, the news and contextual data stream includes the following features: the number of past volatility interruptions in both the overall market and the individual stock over the last hour and the past 24 hours, differentiated by unnecessary and necessary interruptions; a dummy variable indicating whether fast market is active; trade data from the main market and three alternative venues, including the number of trades, total trade volume, and average volume per trade; news data comprising the number of news items, sentiment, relevance, and similarity scores; the number of aggregated order book snapshots per time interval; the relative distance to the approximated static price barrier; message activity, including counts of order submissions, cancellations, and modifications; and the current minute of the observation. A detailed overview of these features, including their calculation methods and informational content, is provided in Table 10 in Appendix A.

The news and contextual data stream is first processed through an LSTM layer to capture temporal dependencies and extract relevant features. This is then followed by a fully connected layer. Unlike the LOB channel, we avoid using a CNN layer for the news and contextual data channel, as there is no rationale to assume meaningful spatial interactions between individual features, such as news sentiment and past volatility interruptions.

The combined model

The combined model architecture integrates the outputs from both data streams. The LOB data channel's feature vector is concatenated with the feature vector from the news and contextual data channel. The resulting combined feature map is then passed through a final fully connected layer, which outputs the probability of the volatility interruption being unnecessary. A schematic illustration of the whole prediction process is given by Figure 14 in Appendix B with a detailed visualization of the entire model architecture provided in Figure 17.

The deep learning approach is expected to outperform the regression models proposed in Section 4 as it leverages not only pre-calculated features but also raw time series data from the market, alternative venues, news, and other contextual factors. Moreover, it enables the model to automatically extract relevant features for the classification task, potentially leading to more accurate predictions. However, a drawback of this approach is the reduced explainability, as the model's internal operations become more complex and less transparent.

Model training and evaluation

For training the model, the dataset was split into training, validation, and test subsets as described in Section 2. The split was conducted on a random basis. Due to the imbalanced distribution of the class labels, the training dataset was oversampled using the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). Oversampling ensures a more balanced training dataset, leading to a more effective training process and potentially better classification performance. The training set was exclusively used for model training. During the training process, the validation set was used to monitor loss and accuracy metrics on a small out-of-sample subset. This monitoring was essential to detect potential overfitting. If overfitting was observed, the training process would have been interrupted and terminated. The test subset, which was never used in the training process to prevent information leakage, was utilized to evaluate the model's performance using the standard classification metrics accuracy, recall, and precision as already discussed in Section 4. Additionally, we consider the F_β-score to assess the balance between precision and recall.

Like in the regression-based approach, the precise identification of unnecessary interruptions is prioritized over general model accuracy. It is more important to detect unnecessary interruptions with high precision to avoid false positives, even if this comes at the expense of lower overall accuracy. Therefore, the deep learning model is also optimized to maximize precision when identifying unnecessary interruptions.

5.2. Results

The model training for the full model converged after 128 iterations. The evolution of the loss function during the training process is documented in Figure 12 in Appendix A. The main objective of this study is to achieve an optimal balance between high precision and recall for predicting unnecessary interruptions, with the primary emphasis on precision to avoid missclassified but relevant interruptions. This balance can be adjusted by testing different cut-off values[29] for the predicted probabilities that indicate unnecessary interruptions.

Figure 8 presents the precision-recall curve, which illustrates the trade-off between precision and recall at different thresholds. This curve is instrumental in identifying an optimal threshold for the underlying prediction task. It becomes evident that thresholds ranging from 0.7 to 0.9 appear to be suitable, given the classification task at hand. While those above 0.7 lead to sufficiently high precision scores, thresholds above 0.75 do further increase precision, which, on the other hand, comes at the cost of a higher decrease in recall. To analyze the impact of the different thresholds within this range on model performance more closely, Table 6 presents the precision and recall scores for thresholds ranging from 0.7 to 0.9. The results indicate that using a threshold of 0.85 or higher yields high precision scores exceeding 84% in identifying unnecessary interruptions. However, this comes at the cost of lower recall scores,

[29] As the output of the model is a continuous value (i.e., the probability of the sample being in the target class), the value needs to be transformed into a binary value representing the class which is done by defining a cut-off value. Every output higher than the cut-off value is considered as part of the target class, in this case, an unnecessary volatility interruption.

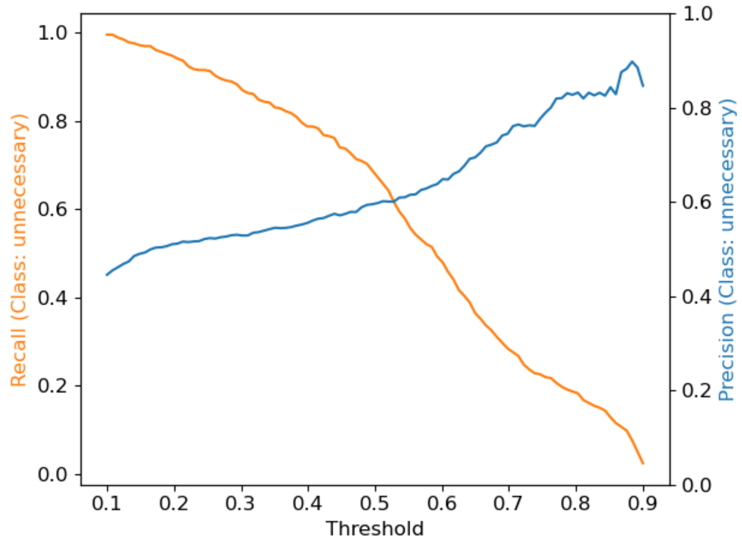

Figure 8: Precision-recall curve of model outputs at different thresholds for labeling as unnecessary This figure shows the achieved precision and recall scores for predicting the target class of unnecessary volatility interruptions at different thresholds. The recall scores are depicted in orange on the left y-axis, while the precision scores are shown in blue on the right y-axis. On the x-axis are different thresholds for the translation of the probabilistic model output to binary results.

with only a relatively small proportion (2% to 13%) of all actually unnecessary interruptions being detected. Conversely, implementing the model with lower thresholds results in significantly higher recall scores. For instance, at a threshold of 0.7, the recall reaches nearly 30%, indicating that nearly one-third of all unnecessary interruptions are correctly identified. However, the precision at this level is lower, just under 75%, indicating that approximately 25% of the interruptions labeled as unnecessary are not part of that class. Larger thresholds are also supported by the F_β-score, which summarizes a model's predictive performance by balancing precision and recall. With a β of 0.1 (0.2), ten (five) times more weight is placed on achieving high precision in predicting unnecessary interruptions, thereby minimizing false positives. This means avoiding predictions of unnecessary interruptions when they are in fact justified. For a β of 0.1, the optimal balance between precision and recall is achieved at a threshold of 0.875, whereas for a β of 0.2, the score reaches its peak at a threshold of 0.775.

We provide a corresponding analysis for the reduced prediction model that relies solely on market data in Appendix C. The precision-recall curve is shown in Figure 20, and detailed classification metrics for various thresholds are reported in Table 11. The results suggest similar threshold behavior in terms of identifying optimal performance; however, overall classification metrics are consistently lower compared to the full model. For instance, at a threshold of 0.875, the reduced model achieves a precision that is 1.5 percentage points lower than the full model, and its recall is only half as large. At a threshold of 0.775, the lite model's precision decreases by 4.5 and its recall by 2 percentage points. In summary, while the reduced model performs slightly worse than the full model, its predictive quality remains reasonably robust.

Furthermore, we conduct a sensitivity analysis on different lengths of the observation window preceding each interruption, which serve as input for the prediction model. The results are presented in Appendix D, with the precision-recall curves for the additional windows of 5, 3, and 1 minutes shown in Figure 21, and detailed classification metrics provided in Table 13. Overall, the predictive performance remains relatively consistent across the tested windows. Shorter windows, i.e., those using more granular market data, generally lead to improved predictive accuracy. However, this trend holds only up to a two-minute window. While the 1-minute window still outperforms the longer ones, its performance declines slightly compared to the two-minute window, suggesting decreasing performance for shorter time frames.

Table 6: Classification metrics of the deep learning prediction model

Classification scores in this table are obtained by predicting the test dataset, where the minority class of unnecessary interruptions represents 41.8% of the observations. We report results for $\beta \in \{0.1, 0.2\}$ for the F_β-score to emphasize the importance of predicting unnecessary volatility interruptions with a high precision to avoid false positives, i.e., predicting an interruption to be unnecessary although it is actually relevant.

Threshold	Accuracy	Precision	Recall	$F_{\beta=0.1}$	$F_{\beta=0.2}$
0.700	0.659	0.743	0.282	0.731	0.699
0.725	0.651	0.760	0.241	0.744	0.702
0.750	0.650	0.782	0.226	0.763	0.714
0.775	0.649	0.825	0.204	0.800	0.738
0.800	0.643	0.825	0.184	0.798	0.728
0.825	0.635	0.830	0.158	0.796	0.713
0.850	0.626	0.845	0.130	0.801	0.698
0.875	0.617	0.882	0.098	0.817	0.674
0.900	0.590	0.846	0.024	0.631	0.364

Given these findings, when aiming for a conservative application of the prediction model, a β of 0.1 should be used, for which our prediction model yields a precision of above 88% at a recall of nearly 10%. With a β of 0.2, which should be picked in cases of less conservative implementations, a precision of above 82% with a recall of 20% is achieved by the proposed model. This offers the potential to prevent the triggering of a volatility interruption when there is a high likelihood that this interruption is unnecessary. Such an approach would minimize delays in price determination and enhance overall market quality by mitigating the negative consequences of volatility interruptions. Our approach can either be fully integrated into a circuit breaker mechanism or serve as a decision support system for market operators to shorten the auction phase if the probability of an unnecessary interruption increases, as more data becomes available during the auction. When implementing this approach in real-world exchange systems, the optimal threshold selection should depend on the level of automation relative to human intervention. In a largely automated environment, a more restrictive mechanism (i.e., higher thresholds) is preferable to minimize false positives. In contrast, when the mechanism is used as a decision support system with a higher level of human supervision, a less restrictive setting may be more suitable tolerating more false positives, as final decisions are reviewed manually by the market supervision team. Accordingly, in settings with high automation, using an F_β-score with a lower β could help identify a suitable threshold, as this is emphasizing precision and stressing the sensitivity to false positives. Conversely, in lower automation settings, a higher β may be more appropriate, as it places greater emphasis on recall, reducing the risk of missed detections. In the following Section 5.3 we further discuss the choice of the decision threshold from a general welfare perspective.

5.3. Welfare implications

To evaluate the practical relevance of the proposed prediction model, we examine its implications for market quality and welfare. While the model is designed to complement existing safeguard mechanisms by predicting whether a volatility interruption triggered according to the rule-based system is necessary, it leads to a trade-off between precision (correctly identifying unnecessary interruptions) and recall (capturing as many unnecessary interruptions as possible). To translate model performance into welfare outcomes, we distinguish between four possible prediction outcomes, each with different implications for market quality and welfare.

First, when the model predicts an interruption to be necessary and the interruption is indeed necessary (a true negative), no change in welfare arises relative to the current rule-based system, since the interruption is triggered under both regimes. Second, if the model incorrectly predicts an interruption to be necessary when it was in fact unnecessary (a false negative), the interruption is still triggered as it would be in the rule-based system, and again no welfare change occurs.

Third, welfare improvements arise in the case of a true positive, where the model correctly identifies an unnecessary interruption and allows continuous trading to proceed. To quantify the welfare gain in such cases, we focus on the increase in transaction costs typically observed following a volatility interruption.[30] Specifically, we estimate the change in the average effective spread from the pre- to the post-interruption period and multiply this difference by the trading volume executed in the post-interruption window. Averaging this across all true positives yields an estimated welfare gain of €1,128.40 per avoided unnecessary interruption, with this case occurring in 7.55% of all interruptions. This calculation is a lower-bound estimate, as it does not account for additional benefits, such as increased order book depth, uninterrupted opportunities for risk sharing (see Chen et al., 2024), and improved price efficiency that would arise from avoiding an unnecessary interruption.

Fourth, cases where the model incorrectly prevents a necessary interruption (false positives) may lead to disorderly market conditions and unsubstantiated price jumps.

[30] As shown in the market quality analysis in Section 2, volatility interruptions are associated with widened bid-ask spreads, reduced order book depth, and increased volatility in the post-interruption phase. This is consistent with the findings of Hautsch and Horvath (2019).

Estimating the welfare loss in these cases is more complex.[31] To approximate this cost, we construct a counterfactual price path by extrapolating the pre-interruption trend[32] and compare it to the price established during the auction of the volatility interruption. The difference between these two prices, multiplied by the auction volume, provides a proxy for the welfare loss from failing to interrupt continuous trading. This yields an estimated loss of €3,017.76 per false positive, which occurs in 1.60% of model predictions.

Aggregating the gains and losses from all four prediction outcomes (for a prediction threshold of 0.775), weighted by their empirical frequencies, and multiplying them by the number of volatility interruptions results in the net average welfare effect when applying the model. Table 7 summarizes the welfare impact per prediction type, along with the associated frequencies and total welfare contribution. Based on the total number of 856[33] volatility interruptions observed in DAX40 stocks during 2024, the estimated annual welfare gain amounts to €31,595. Extending the analysis to all stocks traded on Xetra, which experienced 68,457 interruptions in 2024, the model would generate an estimated total annual gain of €2.53 million, showing that these gains scale meaningfully.[34] If implemented across all asset classes on Xetra or even across other trading venues that employ single-stock circuit breakers, the potential welfare gains would scale further.

We emphasize that these figures are not precise forecasts but serve as back-of-the-envelope estimates intended to approximate the potential welfare impact of the model. The calculations are intentionally conservative, focusing solely on measurable changes in transaction costs. They do not account for additional benefits such as improved order book depth, more efficient risk sharing, or enhanced price efficiency, which would likely amplify the welfare gains from prevented unnecessary interruptions.

To further explore the welfare implications of the model's application, we analyze how different prediction thresholds, which directly affect the trade-off between precision and recall, translate into overall welfare gains associated with the prediction model. Higher thresholds increase the model's precision, reducing the risk of misclas-

[31] Unlike theoretical models such as Chen et al. (2024), empirical analyses are limited by the fact that trader utility is unobservable, and differences in trader types or motives cannot be discerned. Furthermore, because all observed interruptions in the data were triggered, the counterfactual scenario (i.e., what would have happened without the interruption) cannot be directly observed. Following Chen et al. (2024), we adopt a paternalistic planner's perspective, where speculative gains due to price jumps in the absence of a justified interruption are considered inefficient.

[32] To estimate the counterfactual price path, we use a regression model similar to the one employed to identify the dampening effect (see Equation 1), but restrict it to the pre-interruption period and fit it separately for each volatility interruption: $p_i = \beta_0 + \beta_1 t$. Here, p_i denotes the midpoint price observed at each 200-millisecond interval in the pre-interruption phase and t is a time index ranging from 1 to 600. The counterfactual price is then projected at 2 minutes and 15 seconds after the interruption trigger, corresponding to the average duration of a volatility interruption.

[33] In 2024, a total of 856 volatility interruptions occurred in the analyzed DAX40 stocks after applying the data cleaning steps detailed in Table 1, i.e., excluding cases without sufficient continuous trading in the pre-interruption window.

[34] A total of 125,224 volatility interruptions occurred on Xetra in 2024. This includes 68,457 interruptions in equities, 26,994 in ETFs, 23,530 in ETNs, and 6,243 in ETCs.

Table 7: Estimated welfare impact by prediction outcome and aggregate annual gains

This table summarizes the welfare impact of each prediction outcome, including the estimated gain or loss per event, its empirical frequency, and its contribution to average welfare. The results are based on the full model using 2-minute pre- and post-interruption observation windows and a prediction threshold of 0.775. Annual welfare gains are estimated using the observed number of volatility interruptions in DAX40 stocks and across all stocks traded on Xetra in 2024. All values are reported in thousand euros, except for frequencies, which are expressed as percentages.

Prediction Outcome	Welfare Impact per Event	Frequency	Welfare Effect DAX40, 2024 (N = 856)	Welfare Effect all Stocks, 2024 (N = 68,457)
True Negative	0	61.40%	0	0
False Negative	0	29.45%	0	0
True Positive	+1,128.40	7.55%	72.93	5,832.14
False Positive	-3,017.76	1.60%	-41.33	-3,305.39
Net Welfare Gain per Year			**31.60**	**2,526.75**

sifying necessary interruptions as unnecessary, but come at the cost of lower recall. Conversely, lower thresholds raise recall but risk a greater number of false positives.

To quantify this trade-off, we compute the net average welfare gain per interruption for a range of threshold values, taking into account the corresponding precision and recall scores, as well as the estimated welfare gains and losses associated with true positives and false positives. As shown in Figure 9, welfare is maximized at a threshold of 0.775, where precision with 82.5% (see Table 6) is high enough to avoid costly false positives, while recall with 20.4% remains sufficient to avoid a meaningful number of unnecessary interruptions. We also find that the model starts to yield net positive welfare contributions from a threshold of 0.70 onwards. Below this point, the cost of incorrect predictions begins to outweigh the benefits from correctly identified unnecessary interruptions. Above the optimal threshold of 0.775, welfare decreases again due to further decreases in recall.

In summary, our analysis suggests that implementing the prediction model, particularly in a high-precision configuration, can lead to meaningful improvements in market quality and overall welfare, while preserving the existing rule-based system as a fallback in uncertain cases that do not cross the prediction threshold to be classified as unnecessary.

6. Limitations and robustness tests

While this study offers valuable insights into improving the effectiveness of circuit breaker mechanisms in securities markets, certain limitations should be acknowledged when interpreting and applying our findings.

Our analysis primarily examines a specific implementation of circuit breakers - volatility interruptions. Although volatility interruptions are a common safeguard across European stock exchanges, they are less frequently used in other parts of the world (Gomber et al., 2016). In the U.S., circuit breakers are typically implemented as

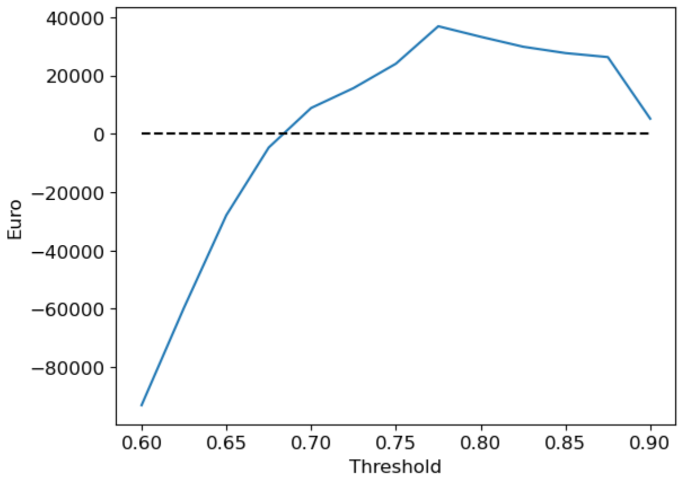

Figure 9: Net economic effect given different prediction thresholds
This figure shows the net welfare effect when applying the prediction model with varying prediction threshold between 0.6 and 0.9, and scaled on an annual basis based on the 856 volatility interruptions observed in the analyzed DAX40 stocks in 2024. The results are based on the full model using 2-minute pre- and post-interruption observation windows. Annual welfare effects are reported in euros.

trading halts, with the Limit Up/Limit Down mechanism governing single-stock trading halts (U.S. Securities and Exchange Commission, 2012). Despite the differences between volatility interruptions and trading halts, both types of circuit breakers are triggered based on pre-determined price thresholds. Our methodology is broadly applicable to rule-based circuit breakers in general, making our results relevant to other mechanisms, including the U.S. Limit Up/Limit Down mechanism. Nonetheless, further comprehensive analyses should be conducted before generalizing our findings to other circuit breaker implementations.

Another limitation is the absence of information on the specific orders that triggered the interruptions in our dataset. This data would provide crucial insights to better assess whether these orders were erroneous or reasonable. With access to the triggering order, we could directly identify misconfigured or error-induced orders for which a interruption is necessary. Additionally, such data could enhance our prediction models by providing valuable information to better identify unnecessary interruptions. The triggering order is not included in public market data feeds. However, as the market operator has knowledge of this order, the inclusion of this information into our proposed approach will likely further increase its performance.

To ensure the robustness of our findings and to account for potential alternative explanations for unnecessary interruptions, we conduct several robustness tests based on the probit regression in Section 4, with results detailed in Appendix D.

Table 15 reports the results of a regression analysis that includes stock fixed effects to control for stock-specific influences. Dummy variables were included for each stock, but the coefficients for these dummies were found to be statistically insignificant, indicating that stock-specific factors do not materially affect the results. The coefficient for the interaction term capturing relevant news remains in the same direction as in the main analysis but loses statistical significance. This suggests that while the underlying relationship holds, it may be sensitive to stock-specific variation. The loss of significance is likely attributable to the limited number of news observations in the dataset, i.e., only 7.6% of volatility interruptions are associated with news events, reducing statistical power, particularly when partitioned across individual stocks. All other coefficients remain consistent with the main results, reinforcing the conclusion that stock-fixed effects do not substantially alter the findings.

As highlighted in Figure 2 and described in Section 2, our dataset was significantly affected by the COVID-19 pandemic. To account for potential pandemic-related effects, we conducted a robustness test incorporating a COVID-19 dummy variable for all samples triggered between February and May 2020. Table 16 presents the results of this analysis, showing that unnecessary interruptions tend to occur less frequently during the market turmoil caused by the COVID-19 pandemic, although this effect is not statistically significant in the full model. In this setup, the coefficient of the news interaction term is also found to be insignificant, while the direction remains the same as in the main analysis. The same applies for the coefficient of the distance to the static barrier in the full model, indicating some sensitivity to the presence of high market-wide volatility. All other effects remained consistent with our main results. This suggests that our approach is applicable across periods of both high and low volatility and is not influenced by market-wide disruptions.

Lastly, we tested for potential year-specific effects by including dummy variables for each year in our observation period in a regression analysis. The results, presented in Table 17, show that, similarly to the COVID-19 dummy, the year 2020 tends to be less associated with unnecessary interruptions, although not being significant in the full model. The only year that shows a significant effect across all models is 2024, also with a negative effect on the likelihood of unnecessary interruptions. This finding is plausible, as 2024 stands out as the only year in our sample without major market-wide disruptions, such as the COVID-19 pandemic or the Russian invasion of Ukraine. In the absence of such events, there is less need for rapid incorporation of market-moving information, which may reduce the occurrence of unnecessary interruptions. Apart from this sensitivity to more stable market periods, no notable deviations from our main results were observed. This suggests that our findings are robust across different years and not driven by year-specific anomalies.

7. Practical Implementation

7.1. Key implementation decisions

This section outlines the key decisions and considerations necessary for implementing our prediction model. Market operators and regulators can use these implementation guidelines, along with the configurable parameters and performance benchmarks outlined in the subsequent section for integration of the model in real-world trading systems. We assume that a rule-based volatility interruption mechanism or a similar single-stock circuit breaker triggered based on deviations from reference prices already exists. Table 8 provides an overview of all implementation decisions.

Choice of observation window and sampling frequency: To ensure applicability in high-frequency markets, we use an observation window of 2 minutes before and after each volatility interruption. Simultaneously, we apply a high sampling frequency of 200 milliseconds, capturing fine price dynamics as relevant for financial markets in a high-frequency trading environment. We also conduct a sensitivity analysis using alternative windows of 1, 3, and 5 minutes with corresponding sampling frequencies of 100, 300, and 500 milliseconds, respectively. These parameters can be further tailored based on trading intensity across different stock segments.

Number of clusters: For the clustering approach, the number of clusters should be sufficiently large to capture the diversity of market dynamics surrounding volatility interruptions. Accordingly, the chosen number should not fall below the threshold identified by the elbow method. While a larger number of clusters can enhance the model's ability to distinguish between different market conditions, it may also reduce interpretability and make visualizations and manual inspection more complex.

Statistical significance thresholds for market quality changes: Our classification method evaluates changes in market quality indicators (such as spread, depth, and volatility) before and after interruptions using hypothesis testing. Thresholds for statistical significance (e.g., 90%, 95%, or 99%) can be selected based on the desired level of conservativeness in identifying unnecessary interruptions.

Dampening effect threshold: We define an interruption as unnecessary only if the price trend continues in the same direction and is not dampened by more than 50%. This 50% dampening threshold is parameterizable and can be adjusted according to what is considered a "sufficient" post-interruption dampening.

Use of external data sources: The full model incorporates external data sources such as news, their sentiment and relevance, and trading activity on alternative venues. However, given potential limitations in real-time data access and processing, we also provide a reduced model that relies solely on internal market data from Xetra. Both variants are evaluated, allowing for flexible adoption based on available infrastructure.

Insufficient pre-interruption data for prediction: In practice, overlapping volatility interruptions or insufficient pre-interruption data do not pose a major challenge for implementation. If the model cannot access a full observation window, e.g., due to proximity to a previous interruption or auction, a conservative fallback approach can be applied, allowing the interruption to proceed under current rules. To operationalize this, the system must track the elapsed time since the last auction or interruption: at least 2 minutes after a scheduled auction or 4 minutes after a prior interruption are required to ensure clean data windows.

Full integration in the matching engine vs. decision support system: The prediction model can be implemented in two distinct operational settings. First, it can be fully integrated into the existing rule-based circuit breaker mechanism, automatically determining whether continuous trading should be interrupted based on the model's classification. Alternatively, it can serve as a decision support system for market supervision. In this semi-automated setting, the model could, for example, assist in terminating interruptions earlier when they are predicted to be unnecessary. This mode offers greater discretion and comes with less stringent latency requirements.

Table 8: Overview of configurable implementation parameters

This table provides an overview of configurable parameters and model choices relevant for implementing the prediction framework in a live trading environment.

Parameter / Model Choice	**Options / Values**
Observation window	1, 2, 3, or 5 minutes
Sampling frequency	100ms, 200ms, 300ms, or 500ms
Number of clusters	12 (default, adjustable)
Significance threshold for market quality changes	90%, 95%, or 99%
Post-interruption dampening threshold	50% (adjustable between 0% and 100%)
External data inputs	None / News / Alternative venue data
Insufficient pre-interruption data	Trigger interruption as in rule-based system
Deployment mode	Full integration / Decision support system

7.2. Computational requirements

While data aggregation, feature calculation, and model training is computationally intensive and takes several hours depending on system architecture (see Table 9), the prediction latency is extremely low. On our infrastructure, the CNN-based model generates a prediction in approximately 27 milliseconds.[35] Given that trades in DAX40 stocks occur on average every 20 seconds, and that the executable order triggering the interruption is not included in the input data, the model has ample time to compute its prediction. Furthermore, it can process predictions incrementally with each incoming order book update, making inference latency non-critical in practical settings. In latency-optimized trading environments, further reductions in inference time are achievable by the respective market operator.

Table 9: Approximate computation times for model components

This table reports the approximate runtimes for the main components of the model pipeline. The average inference time (the most latency-sensitive element for practical deployment) is based on 100 runs, with the corresponding standard deviation included in parenthesis to reflect runtime variability.

Component	**Approximate Runtime**
Data aggregation	24 hours
Feature calculation	3 hours
Normalization	2 minutes
Clustering (training)	2 hours
Prediction model training	1 hour
Prediction / inference	27 ms (SD = 4.81 ms)

[35] System specifications: 2x Nvidia A100 40GB, 256 GB RAM, AMD EPYC 74F3 3.2 GHZ 24-Core Processor.

8. Conclusion

Our study addresses a significant gap in the literature on circuit breakers, as it demonstrates that it is possible to identify and avoid unnecessary trading interruptions. Interruptions caused by circuit breakers can negatively impact market quality by delaying price discovery and disrupting liquidity, leading to a trade-off between their protective role and adverse effects on market quality (Hautsch and Horvath, 2019). Our research presents a novel approach based on machine learning techniques to identify unnecessary volatility interruptions that delay price discovery because they are triggered within an ongoing price trend before and after the interruption and are characterized by reductions in market quality. Based on this identification, we further develop a deep learning model that is able to predict unnecessary volatility interruptions using ex-ante order book information, news, and other contextual features. This approach enables a more nuanced application of circuit breakers, facilitating the development of advanced market safeguards that are only triggered when necessary. Therefore, our study adds to the discussion of improved circuit breaker mechanisms, such as the forward-looking circuit breaker proposed by Bongaerts et al. (2024). Our study also contributes to the broader literature on market microstructure by applying advanced machine learning techniques to a critical market design aspect.

By analyzing the circumstances in which unnecessary volatility interruptions are triggered, we find that they are more likely to occur when liquidity is high and when relevant news are present. Large price fluctuations observed under these conditions rather point to well-functioning price discovery instead of erroneous price jumps. Moreover, volatility interruptions are also more likely to be unnecessary when the last price prior to the interruption is near the triggering threshold. This indicates that the existing simplistic rule-based mechanism is not capable of differentiating between plausible price changes and unsubstantiated price jumps.

The practical implications of our findings are considerable. Exchange operators can utilize our predictive model to enhance their safeguard mechanisms by reducing the frequency or duration of unnecessary interruptions, thereby improving overall market efficiency. A welfare analysis confirms that implementing the model in practice leads to a net gain in welfare. Moreover, the model offers a novel approach to mitigating the magnet effect associated with rule-based circuit breakers (Chen et al., 2024). Since interruptions are only triggered if not predicted to be unnecessary, the approach decouples the direct link between price proximity to the threshold and the initiation of an interruption. This softening of the rule-based mechanism can reduce the incentive for traders to anticipate or accelerate toward interruptions, thereby dampening the self-reinforcing dynamics associated with the magnet effect that typically exacerbate market volatility. Additionally, our results can inform regulatory discussions, potentially leading to more nuanced and effective rules for circuit breakers.

Future research can apply our methodology to different markets and types of circuit breakers, further validating the robustness of our findings. Additionally, exploring other machine learning techniques and incorporating non-public information from market operators will likely enhance the predictive accuracy of our models.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4o in the writing process in order to check grammar and improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Abad, D., Pascual, R., 2007. On the magnet effect of price limits. European Financial Management 13(5), 833–852.

Abad, D., Pascual, R., 2010. Switching to a temporary call auction in times of high uncertainty. Journal of Financial Research 33(1), 45–75.

Abad, D., Pascual, R., 2013. Holding back volatility: Circuit breakers, price limits, and trading halts, in: Market Microstructure in Emerging and Developed Markets: Price Discovery, Information Flows, and Transaction Costs. John Wiley & Sons. chapter 17, pp. 303–324.

Bongaerts, D., Draus-De Luca, S., Van Achter, M., 2024. Circuit breakers and market runs. Review of Finance (forthcoming) .

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16, 321–357.

Chen, H., Petukhov, A., Wang, J., Xing, H., 2024. The dark side of circuit breakers. The Journal of Finance 79(2), 1405–1455.

Cho, D.D., Russell, J., Tiao, G.C., Tsay, R., 2003. The magnet effect of price limits: Evidence from high-frequency data on Taiwan Stock Exchange. Journal of Empirical Finance 10(1-2), 133–168.

Clapham, B., Gomber, P., Haferkorn, M., Panz, S., 2017a. Managing excess volatility: Design and effectiveness of circuit breakers. Working Paper .

Clapham, B., Gomber, P., Panz, S., 2017b. Coordination of circuit breakers? Volume migration and volatility spillover in fragmented markets. Working Paper .

Deutsche Boerse Group, 2024. T7 Release 12.1: Market model for the trading venue Xetra. URL: https://www.xetra.com/resource/blob/3873542/fd4c485e86257fd9d7320ded6a6c4df3/data/T7_Release_12.1_-_Market_Model%20_Xetra.pdf.

Easley, D., López de Prado, M., O'Hara, M., Zhang, Z., 2021. Microstructure in the machine age. The Review of Financial Studies 34(7), 3316–3363.

European Commission, 2014. Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on Markets in Financial Instruments and amending Directive 2002/92/EC and Directive 2011/61/EU Text with EEA relevance: MiFID II. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0065.

European Securities and Markets Authority, 2022. ESMA report on trends, risks and vulnerabilities, No. 2, 2022. URL: https://www.esma.europa.eu/sites/default/files/library/esma50-165-2229_trv_2-22.pdf.

Fama, E.F., 1988. Perspectives on October 1987, or, what did we learn from the crash? Center for Research in Security Prices, Graduate School of Business, University of Chicago.

Gomber, P., Clapham, B., Haferkorn, M., Panz, S., Jentsch, P., 2016. Ensuring market integrity and stability: Circuit breakers on international trading venues. The Journal of Trading 12(1), 42–54.

Gomber, P., Haferkorn, M., Lutat, M., Zimmermann, K., 2013. The effect of single-stock circuit breakers on the quality of fragmented markets, in: Rabhi, F.A., Gomber, P. (Eds.), Lecture Notes in Business Information Processing, Vol. 135, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 71–87.

Hautsch, N., Horvath, A., 2019. How effective are trading pauses? Journal of Financial Economics 131(2), 378–403.

Kim, K.A., Rhee, S.G., 1997. Price limit performance: Evidence from the Tokyo Stock Exchange. Journal of Finance 52(2), 885–901.

Kwan, A., Philip, R., Shkilko, A., 2021. The conduits of price discovery: A machine learning approach. Working Paper .

Lauterbach, B., Ben-Zion, U., 1993. Stock market crashes and the performance of circuit breakers: Empirical evidence. The Journal of Finance 48(5), 1909–1925.

Lee, C.M., Ready, M.J., Seguin, P.J., 1994. Volume, volatility, and New York Stock Exchange trading halts. The Journal of Finance 49(1), 183–214.

Lehmann, B.N., 2019. Commentary: Volatility, price resolution, and the effectiveness of price limits. Journal of Financial Services Research 3(2-3), 205–209.

Ma, C.K., Rao, R.P., Sears, S., 1989. Limit moves and price resolution: The case of the treasury bond futures market. The Journal of Futures Markets 9(4), 321–335.

Moise, C.E., 2025. Circuit breakers, illiquidity, and the COVID-19 crisis. Working Paper .

Nagel, S., 2012. Evaporating liquidity. The Review of Financial Studies 25(7), 2005–2039.

O'Hara, M., 2015. High frequency market microstructure. Journal of Financial Economics 116(2), 257–270.

Sifat, I.M., Mohamad, A., 2019. Circuit breakers as market stability levers: A survey of research, praxis, and challenges. International Journal of Finance & Economics 24(3), 1130–1169.

Sirignano, J., Cont, R., 2019. Universal features of price formation in financial markets: Perspectives from deep learning. Quantitative Finance 19(9), 1449–1459.

Subrahmanyam, A., 1994. Circuit breakers and market volatility: A theoretical perspective. The Journal of Finance 49(1), 237–254.

Subrahmanyam, A., 2013. Algorithmic trading, the flash crash, and coordinated circuit breakers. Borsa Istanbul Review 13(3), 4–9.

Syakur, M.A., Khotimah, B.K., Rochman, E., Satoto, B.D., 2018. Integration k-means clustering method and elbow method for identification of the best customer profile cluster, in: IOP Conference Series: Materials Science and Engineering, pp. 1–6. Article No. 012017.

Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18(4), 267–276.

U.S. Securities and Exchange Commission, 2012. Investor bulletin: Measures to address market volatility. URL: https://www.sec.gov/resources-for-investors/investor-alerts-bulletins/investoralertscircuitbreakersbulletinhtm.

U.S. Securities and Exchange Commission, 2016. Opening remarks at the equity market structure advisory committee meeting. URL: https://www.sec.gov/newsroom/speeches-statements/white-remarks-emsac-meeting-080216.

World Federation of Exchanges, 2008. Questionnaire on circuit-breakers for Q4 2008. URL: https://www.world-exchanges.org/storage/app/media/research/Studies_Reports/2008%20WFE%20Circuit%20Breakers%20Report.pdf.

Zhang, Z., Zohren, S., Roberts, S., 2019. DeepLob: Deep convolutional neural networks for limit order books. IEEE Transactions on Signal Processing 67(11), 3001–3012.

Appendix A: Variable description and additional descriptive statistics

This section provides a description of all variables used and their respective calculations in Table 10. Additional descriptive statistics on the approximated static price limits are shown in Figure 10. The number of volatility interruptions per stock during the observation period is illustrated in Figure 11 and the evolution of the loss function during training of the prediction model is depicted in Figure 12.

Table 10: Detailed explanation of all variables used in the models in Section 4 and Section 5

The first column lists the variable names. The second column indicates in which models each variable is used in: P1 refers to the *market quality* model, P2 to the *news and contextual factors* model, and P3 to the *full* model from the probit regressions discussed in Section 4.1. MQ denotes the regression model to analyze the implications of unnecessary interruptions for market quality applied in Section 4.3, and DL refers to the deep learning model described in Section 5. The third column provides a description of each variable and its calculation method. The final column details how the 200-millisecond intervals are aggregated into a single value. For the probit regressions, we apply exponential weighting, whereas the MQ regression in Section 4.3 employs time-weighted aggregation. In this column, "none" refers to no aggregation method as the calculation of the measure results in a single value rather than a time series (e.g., volatility or number of news in the past ten minutes) and "last" means the last observation prior to the interruption is used. In the DL model, the variables are not further aggregated across the time intervals but used as a time series, also denoted as "none".

Feature	Model	Calculation	Aggregation method
rel_spread	P1, P3, MQ	Measures the difference between the best bid and ask prices relative to the midpoint.	Mean
level_1_depth	P1, P3, MQ	Sum of the quoted Euro volume at the first five order book levels.	Sum
trade_volume	P1, P3, MQ, DL	Euro volume of all executed trades.	Sum (DL: none)
midpoint_return_vola	P1, P3, MQ	Measures overall volatility by calculating the standard deviation of midpoint returns. Returns are based on the respective sampling frequency.	None
message_count	P1, P3, MQ	Number of all messages (submissions, deletions, modifications).	Sum
news_count	P2, P3, MQ, DL	Number of news articles mentioning the corresponding company in the last ten minutes.	Sum (DL: none)
news_relevance	P2, P3, MQ, DL	Average relevance score of news articles mentioning the corresponding company in the last ten minutes. Relevance is defined by how prominently the company is referenced in the article, such as in the headline or concluding sentence.	Mean (DL: none)

continued ...

. . . continued

Feature	Model	Explanation	Aggregation method
unnecessary_-interruptions	P2, P3, MQ, DL	Measures the number of triggered volatility interruptions across all DAX40 constituents in the last hour being labeled unnecessary. In the deep learning model, we additionally calculate this measure based on the last 24 hours and for the respective stock.	Last (DL: none)
necessary_-interruptions	P2, P3, MQ, DL	Measures the number of triggered volatility interruptions across all DAX40 constituents in the last hour not being labeled unnecessary. In the deep learning model, we additionally calculate this measure based on the last 24 hours and for the respective stock.	Last (DL: none)
fast_market_dummy	P2, P3, MQ, DL	Dummy variable whether the fast market indicator is set by the market operator or not.	None
distance_to_static_barrier	P2, P3, MQ, DL	Distance of the current price to the approximated static price range. The value is bounded between zero and one with one being close to the barrier and zero being further away.	Last (DL: none)
trade_count_alt_venues	P2, P3, MQ, DL	Measures the overall number of trades executed across the three alternative trading venues (Aquis, Turqoise, Cboe). In the deep learning model is this measure split into three variables each representing one exchange.	Sum (DL: none)
count_trades	DL	Number of executed trades.	None
count_buys	DL	Number of buyer initiated trades.	None
count_sells	DL	Number of seller initiated trades.	None
avg_trade_volume	DL	Average trade size in Euro.	None
news_sentiment	DL	Average sentiment score of news articles mentioning the corresponding company in the last ten minutes.	None
news_similarity	DL	Measures the average similarity of news articles mentioning the corresponding company in the last ten minutes to other older news articles.	None

continued . . .

. . . continued

Feature	Model	Explanation	Aggregation method
count_add	DL	Number of messages representing the adding of a new order.	None
count_delete	DL	Number of messages representing the deletion of a persisting order.	None
count_modify	DL	Number of messages representing the modification of a persisting order.	None
ob_changes	DL	Number of order book updates.	None
minute_of_hour	DL	Number representing the current minute in the time series. Bounded between zero and 59.	None
last_auction_price_return	DL	Return of the observed price and the last auction price.	None

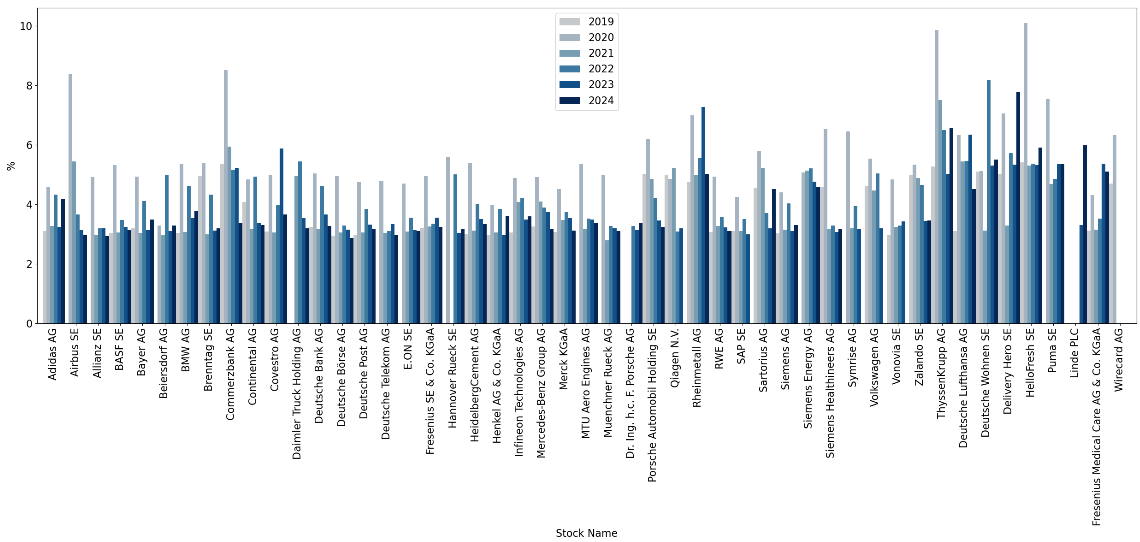

Figure 10: Approximated static price ranges for all DAX40 constituents
This figure shows the approximated static price ranges based on the largest price deviation seen from the last auction price, which serves as reference price for the static threshold triggering the interruption. For brevity and better visualization, only the yearly approximated thresholds are shown, although monthly approximations are used in the models. Missing bars are due to the stock's later listing (e.g., Daimler Truck Holding AG emerged from a spin-off and was first traded in December 2021).

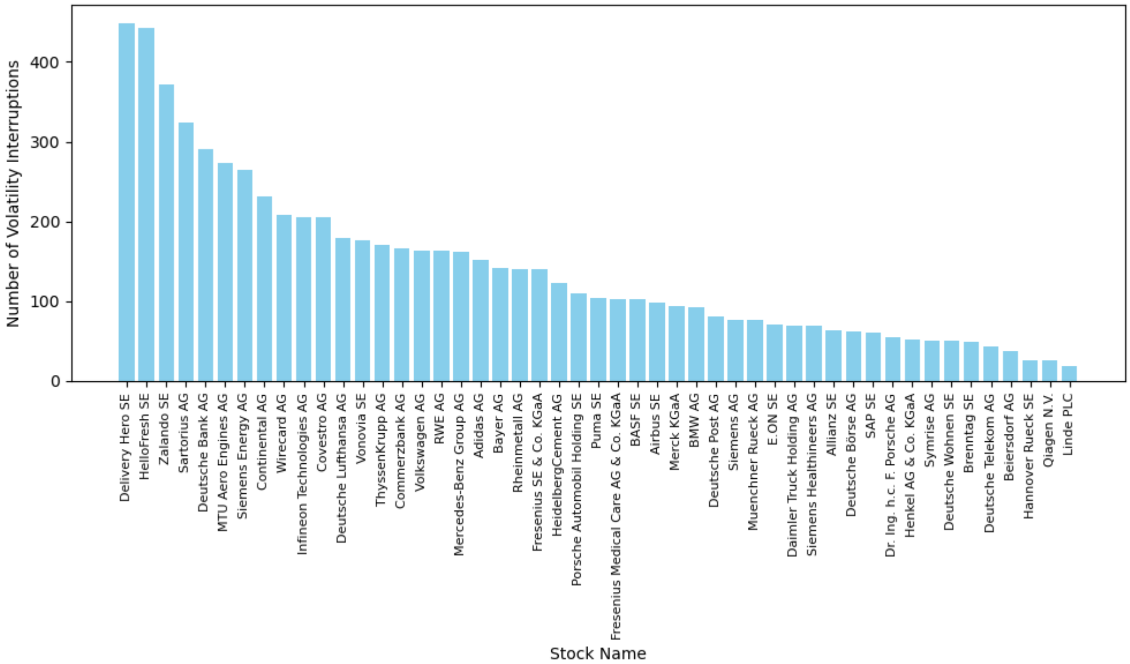

Figure 11: Total number of volatility interruptions for each DAX stock between 2019 and 2024
This figure shows the number of volatility interruptions for each DAX stock from 2019 to 2024. We consider each stock that was part in the DAX during this period as part of our sample. The count of volatility interruptions in this figure is prior to any data cleaning processes.

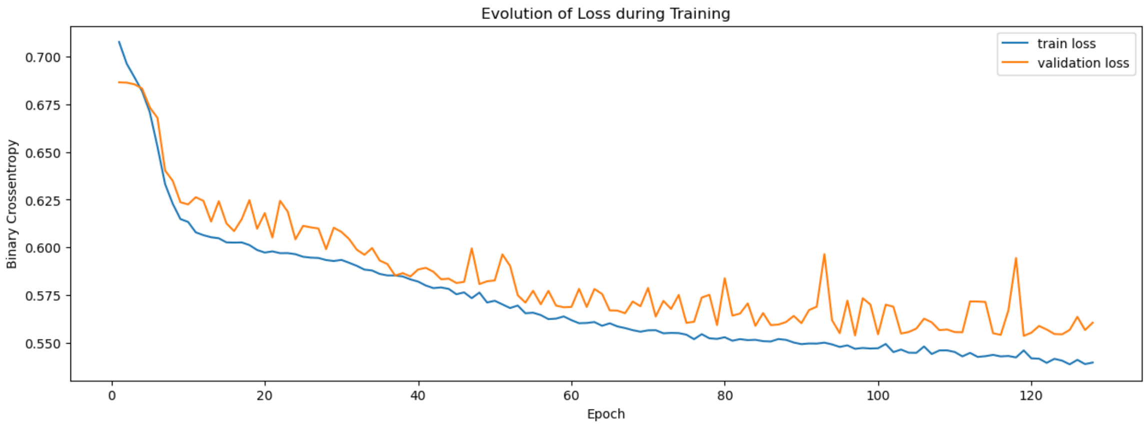

Figure 12: Loss during the training of the classification model for both the training and testing subset
This figure depicts the evolution of the loss function during model training. On the y-axis, the loss, calculated as the binary crossentropy, is shown. The x-axis shows the training iteration called epoch. The blue line depicts the loss for the training dataset and the orange one the loss for the validation dataset.

Appendix B: Architectures of the deep learning models

This section provides more detailed insights into the deep learning model architectures and the whole process on how cluster labels and label predictions for unnecessary volatility interruptions are obtained. In Figure 13 and Figure 14, schematic visualizations are depicted, describing the process of generating cluster labels and the prediction of a volatility interruption being unnecessary or not. In Figure 15, the encoder, and in Figure 16, the decoder architecture of the autoencoder model from Section 3 are shown in detail. Figure 17 shows the detailed architecture of the prediction model from Section 5.

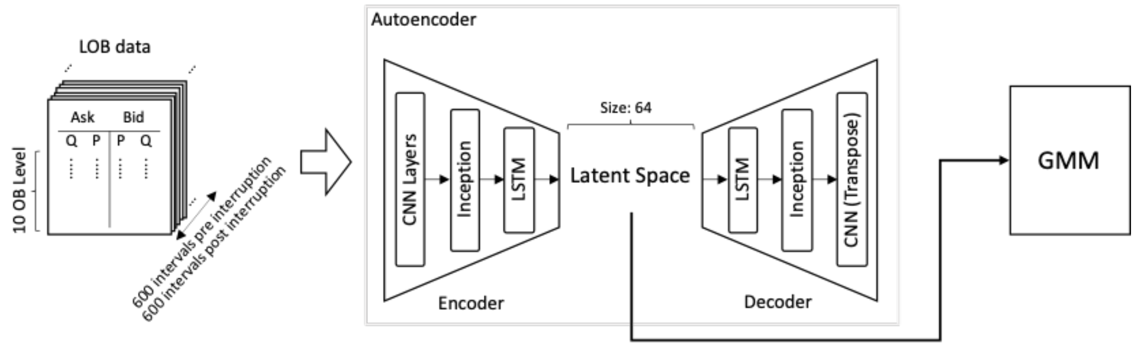

Figure 13: Schematic visualization of the clustering pipeline
This figure shows a schematic visualization of the whole pipeline to obtain cluster assignments. On the left-hand side, the input data is shown, which is a time series of order book snapshots. The middle part illustrates the autoencoder architecture. On the right-hand side, the GMM is contemplated.

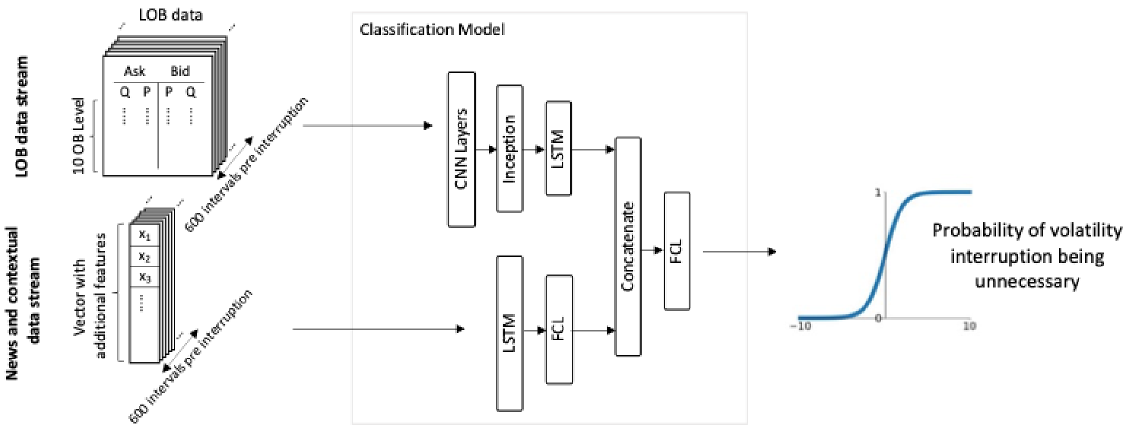

Figure 14: Schematic visualization of the prediction pipeline
In this figure, the whole pipeline to obtain label predictions for unnecessary volatility interruptions is visualized. On the left-hand side, the two input data streams are illustrated: the LOB data stream and the news and contextual data stream. In the middle, the deep learning model architecture is described, combining the two data streams. On the right-hand side, the final probability output using a sigmoid function is shown.

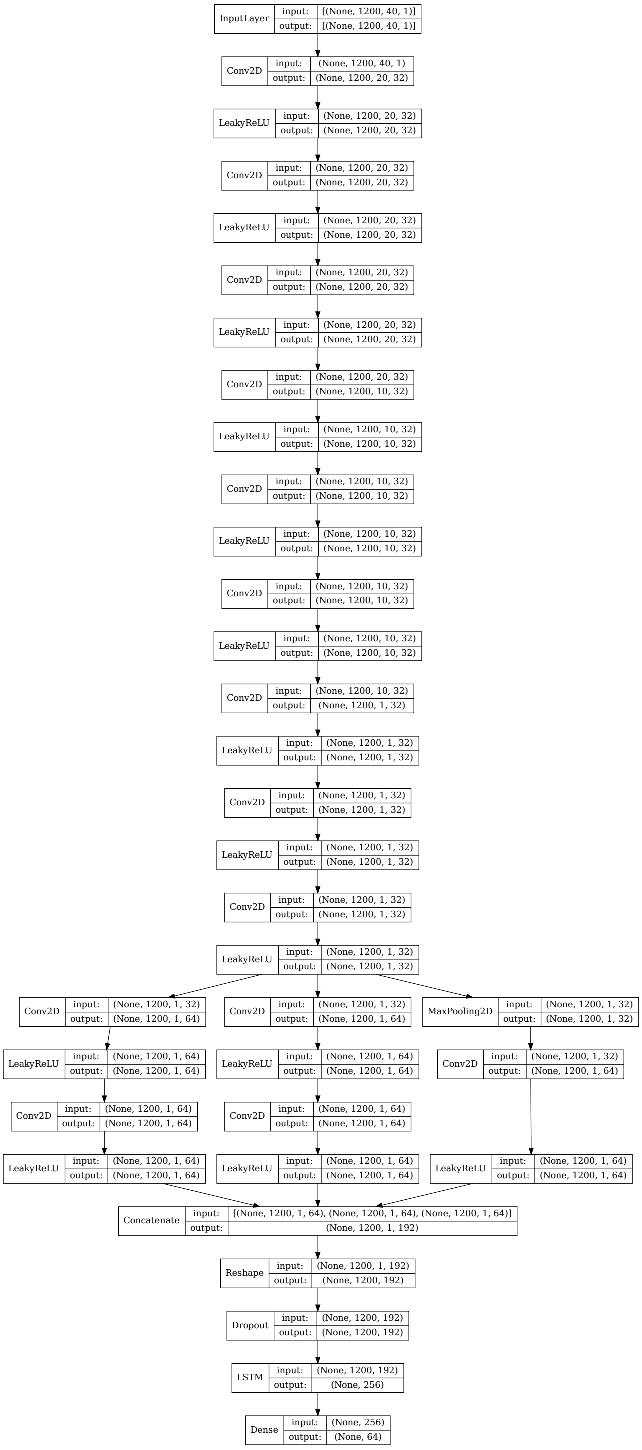

Figure 15: Architecture of the implemented encoder part of the autoencoder model

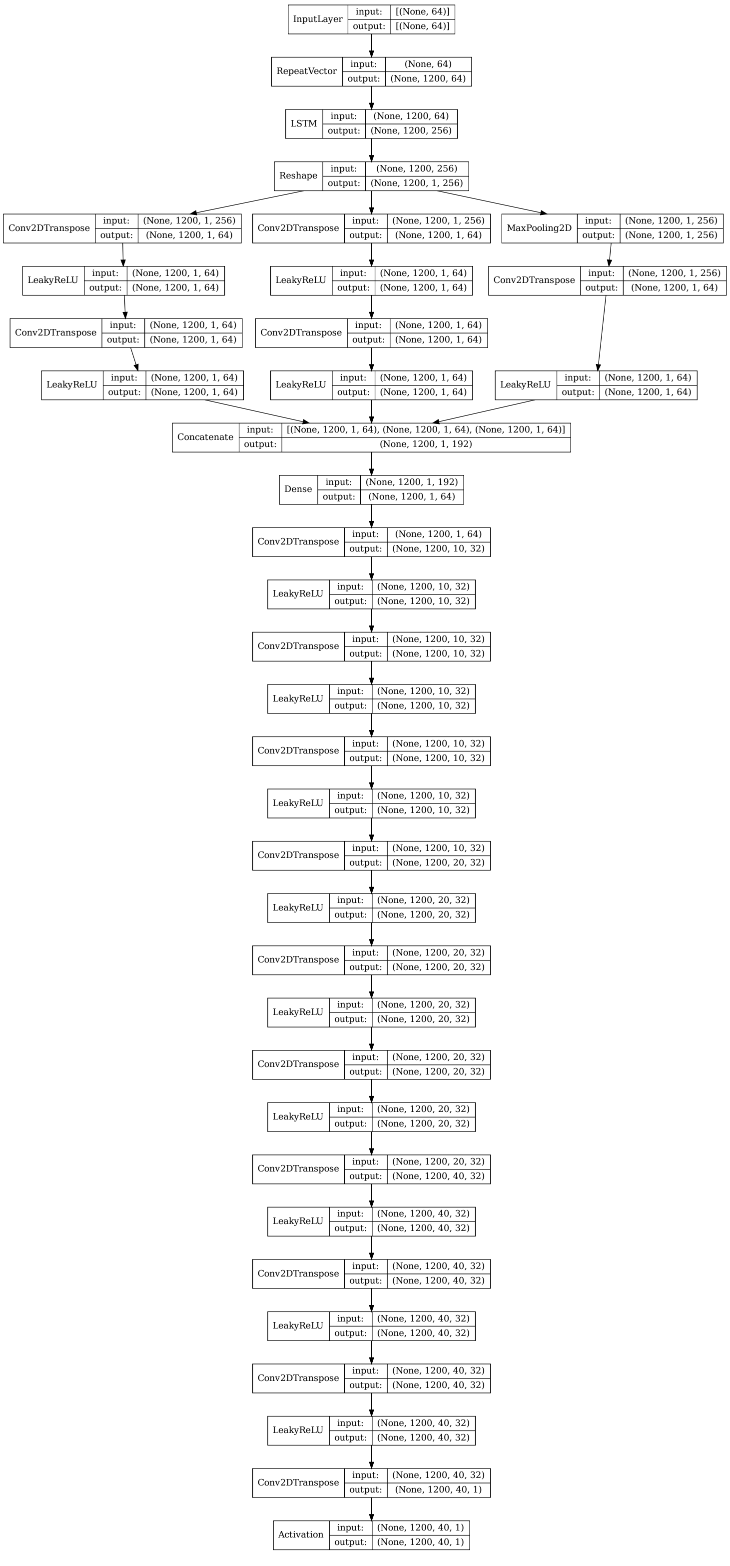

Figure 16: Architecture of the implemented decoder part of the autoencoder model

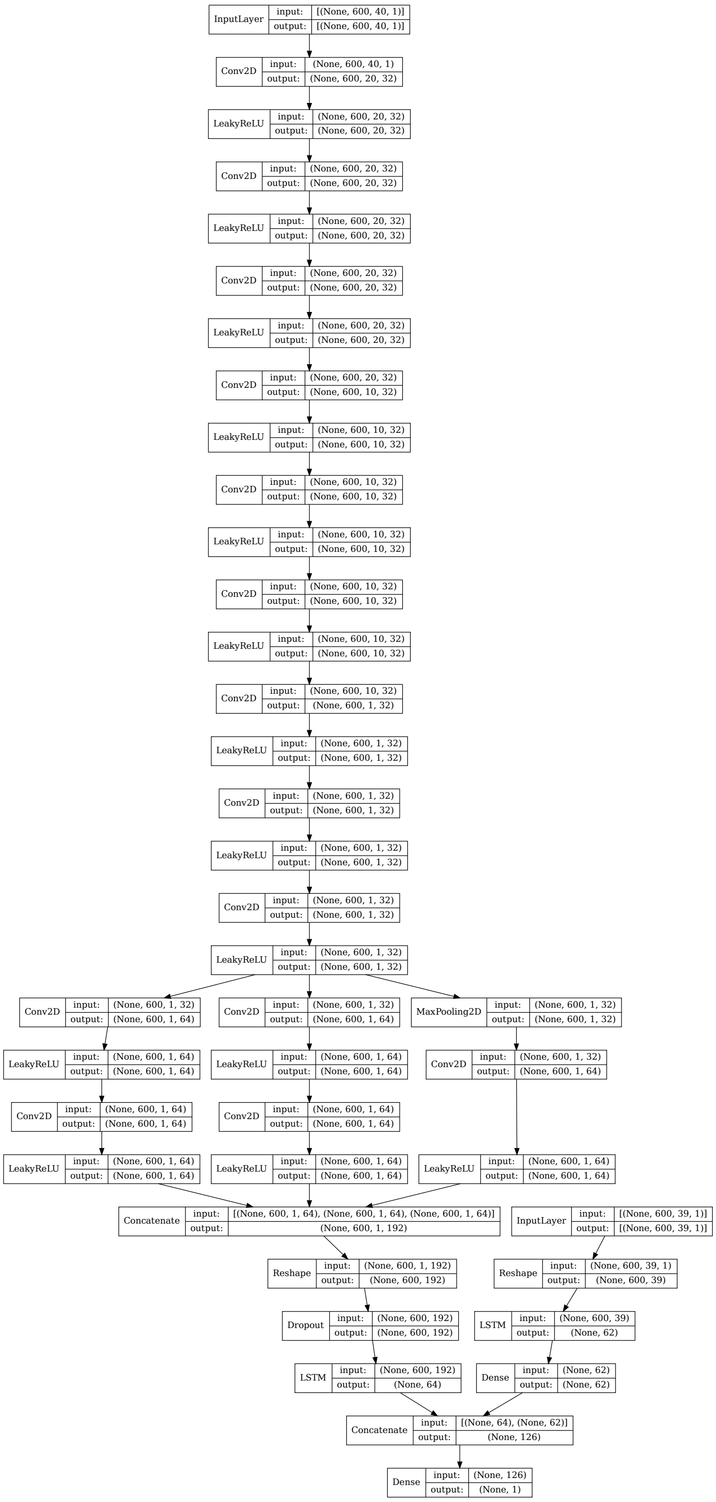

Figure 17: Architecture of the implemented classification model

Appendix C: Additional results

Choice of the number of clusters and cluster probabilities

This section shows the results of the elbow method applied in order to evaluate the choice of the number of cluster k as in Section 3. The results are displayed in Figure 18. Further, we provide insights on the probability distribution of cluster assignments in Figure 19.

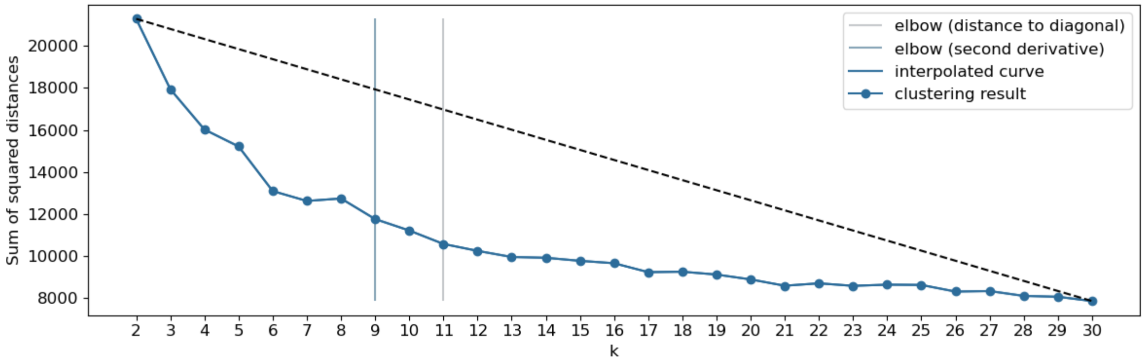

Figure 18: Result of the elbow method
This figure displays the sum of squared distances from all observations to their respective cluster centers, computed from GMM clustering for various values of k. The blue curve represents an interpolation of these values to visualize the overall trend. The dashed gray line denotes the diagonal connecting the endpoints at $k = 2$ and $k = 30$. The two vertical lines indicate the optimal number of clusters as identified by two variants of the elbow method: one based on the maximum second derivative (curvature) of the curve, and the other based on the point with the greatest perpendicular distance to the diagonal.

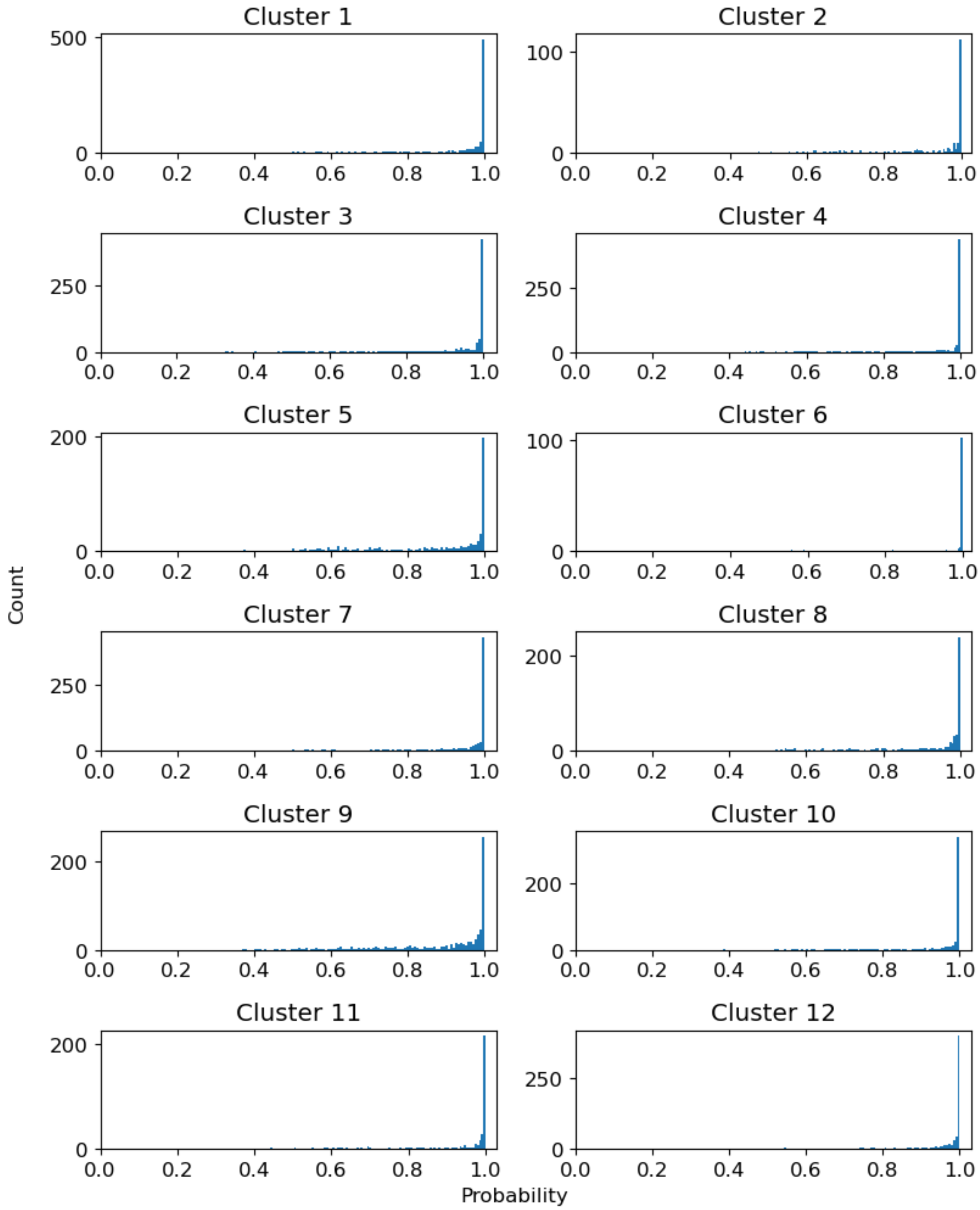

Figure 19: Distribution of cluster probabilities for each cluster
This figure presents, for each cluster, a histogram of the probabilities of observations assigned to the specific cluster. The number of observations in the respective probability bucket is depicted on the y-axis and the probability on the x-axis.

The reduced deep learning prediction model

This section presents the prediction results of the reduced deep learning prediction model. The reduced model solely relies on market data as the input data. In Figure 20, we present the resulting precision-recall curve, and in Table 11, a more detailed overview of achieved classification results given different prediction thresholds.

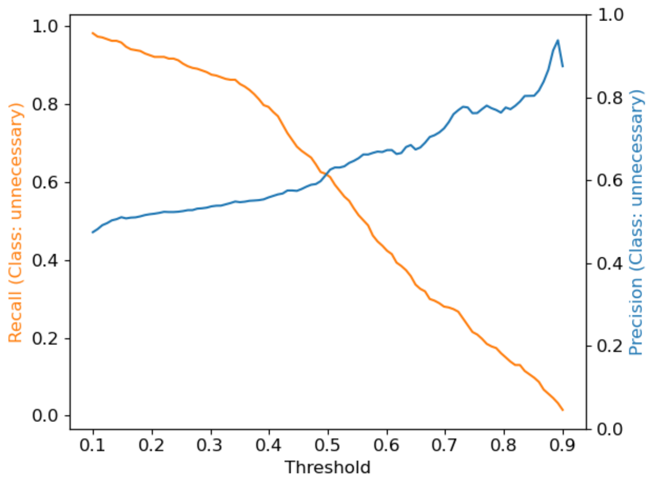

Figure 20: Precision-recall curve of the model outputs at different thresholds for labeling as unnecessary in case of the reduced model
This figure shows the achieved precision and recall scores for predicting the target class of unnecessary volatility interruptions at different thresholds. The recall scores are depicted in orange on the left y-axis, while the precision scores are shown in blue on the right y-axis. On the x-axis are different thresholds for the translation of the probabilistic model output to binary results.

Table 11: Classification metrics of the reduced deep learning prediction model

Classification scores in this table are obtained by predicting the test dataset, where the minority class of unnecessary interruptions represents 41.8% of the observations. We report results for $\beta \in \{0.1, 0.2\}$ for the F_β-score to emphasize the importance of predicting unnecessary volatility interruptions with a high precision to avoid false positives, i.e., predicting an interruption to be unnecessary although it is actually relevant.

Threshold	Accuracy	Precision	Recall	$F_{\beta=0.1}$	$F_{\beta=0.2}$
0.700	0.655	0.725	0.280	0.713	0.683
0.725	0.659	0.771	0.262	0.756	0.717
0.750	0.644	0.766	0.213	0.746	0.696
0.775	0.637	0.780	0.184	0.756	0.694
0.800	0.627	0.772	0.154	0.742	0.669
0.825	0.622	0.789	0.130	0.752	0.661
0.850	0.615	0.810	0.102	0.758	0.639
0.875	0.602	0.867	0.056	0.759	0.558
0.900	0.587	0.875	0.015	0.561	0.275

Table 12: Impact of unnecessary interruptions on market quality

This table reports the full results of the regression specified in Equation 4, which tests whether unnecessary volatility interruptions affect market quality differently from interruptions classified as necessary. The dependent variables include Δrel_spread, the change in the time-weighted average relative spread, $\Delta level\_1\_depth$, the change in the time-weighted average depth at the top of the order book, and $\Delta return\_vola$, the change in volatility measured as the standard deviation of midpoint returns over 200-millisecond intervals. Each variable is computed as the difference between the post- and pre-interruption phases ($Post - Pre$). The row labeled $Average\Delta$ reports the unconditional average of these differences across all volatility interruptions. The key explanatory variable, *unnecessary*, is a binary indicator equal to 1 if the interruption was classified as unnecessary. The control variables mirror those used in the full probit model reported in Table 4, but are calculated using standard time-weighted averages (rather than exponential averages) to ensure consistency with the dependent variable definitions. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

	Δrel_spread	$\Delta level\_1\_depth$	$\Delta return\_vola$
$Average\Delta$	1.27bps	−0.08(€100k)	0.32bps
constant	3.6333***	0.2615***	0.6777***
	(0.3536)	(0.0211)	(0.0590)
unnecessary	−0.3268***	0.0434***	−0.0588***
	(0.1168)	(0.0070)	(0.0195)
rel_spread	−0.1058***	−0.0013***	0.0239***
	(0.0082)	(0.0005)	(0.0014)
level_1_depth	0.0000***	0.0000***	0.0000**
	(0.0000)	(0.0000)	(0.0000)
midpoint_return_vola	0.8623***	−0.0181***	−0.3362***
	(0.0645)	(0.0039)	(0.0108)
trade_volume	0.0007*	0.0004***	0.0002***
	(0.0003)	(0.0000)	(0.0001)
message_count	−0.0237***	−0.0008***	0.0050***
	(0.0021)	(0.0001)	(0.0003)
news_count	0.2627***	−0.0031	0.0672***
	(0.0826)	(0.0049)	(0.0138)
news_relevance	−0.0005*	0.0000	0.0002***
	(0.0003)	(0.0000)	(0.0000)
news_relevance_interaction	−0.0005***	0.0000	−0.0002***
	(0.0002)	(0.0000)	(0.0000)
unecessary_interruptions	−0.0004***	0.0000	0.0000***
	(0.0001)	(0.0000)	(0.0000)
necessary_interruptions	0.0002***	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
fast_market_dummy	0.0025***	0.0000	0.0000
	(0.0004)	(0.0000)	(0.0001)
distance_to_static_barrier	−1.9209***	0.0156	−0.4375***
	(0.3477)	(0.0208)	(0.0580)
trade_count_alt_venues	0.0003	−0.0001***	0.0001**
	(0.0003)	(0.0000)	(0.0001)
R-squared	0.0730	0.5704	0.1430
R-squared Adj.	0.0711	0.5696	0.1413
Observations	6887	6887	6887

Appendix D: Robustness tests and sensitivity analysis

Sensitivity analysis regarding observation windows and sampling frequencies

This section shows the results for the prediction of unnecessary volatility interruptions based on different observation windows prior to the interruption. In Figure 21, we present the precision-recall curves of prediction models using 5-, 3-, 2-, and 1-minute observation windows prior to the interruption. The sampling frequency changes accordingly to the used observation window. To be precise, data is sampled on 500-, 300-, 200-, and 100-millisecond frequencies for the 5-, 3-, 2-, and 1-minute observation windows. In Table 13 a more detailed and comparable view on the prediction results is shown by reporting different classification metrics, namely the accuracy, precision, recall, and the precision at recall score at a minimum recall of 0.2 and 0.1.

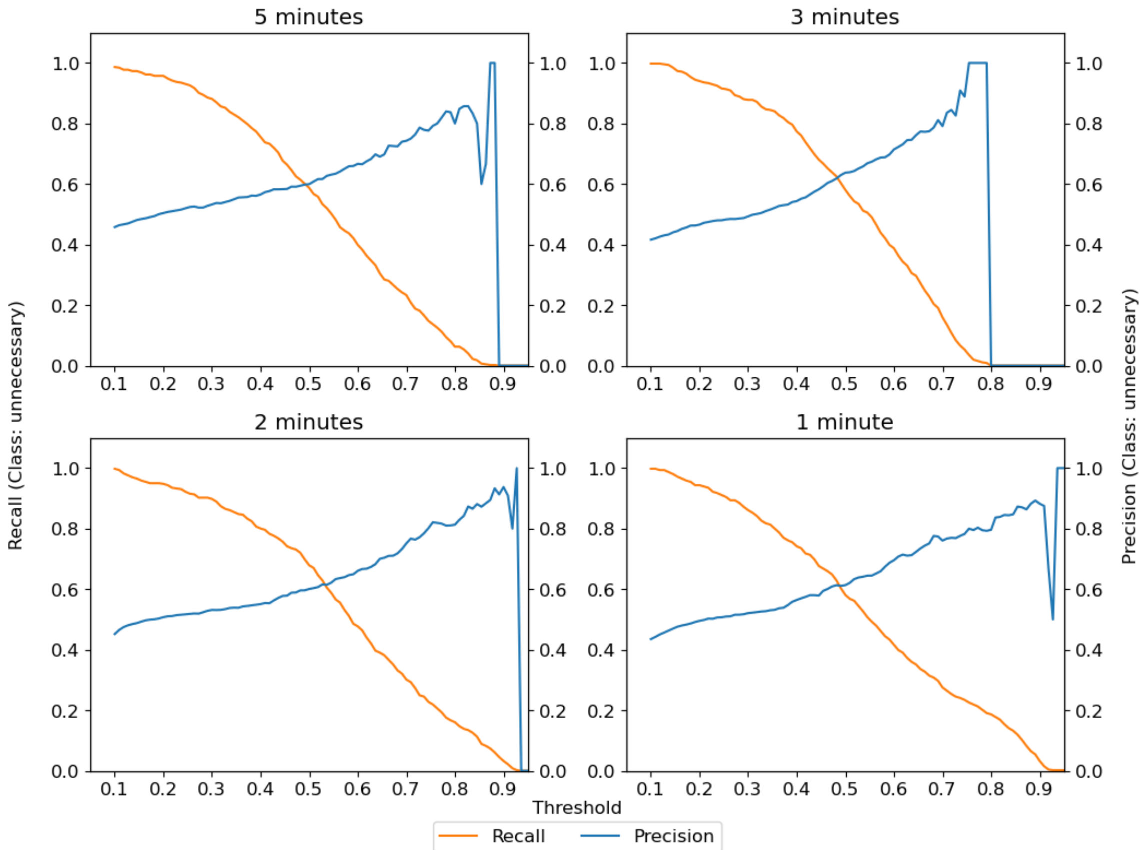

Figure 21: Precision-recall curve for prediction models given different preceding observation windows

This figure shows the achieved precision and recall scores for predicting the target class of unnecessary volatility interruptions at different thresholds for different observation windows. The recall scores are depicted in orange on the left y-axis, while the precision scores are shown in blue on the right y-axis. On the x-axis are different thresholds for the translation of the probabilistic model output into binary results. Each single figure shows the achieved results for the deep learning prediction, given a specific length of preceding information prior to the triggering of the volatility interruption.

Table 13: Classification metrics of the prediction models given different preceding observation windows

This table shows classification results for the deep learning prediction model given different lengths of the preceding observation window. The accuracy score, precision score, and recall score is calculated at the standard threshold of 0.5. Precision and recall are given for the target class of unnecessary volatility interruptions. The precision at recall metric states the maximum achievable precision score while attaining a minimum recall. In this table, this score is given for a minimum recall of 0.1 and 0.2.

Window (minutes)	Accuracy	Precision	Recall	Precision at recall 20	Precision at recall 10
5	0.66	0.60	0.59	0.75	0.82
3	0.69	0.64	0.58	0.79	0.84
2	0.68	0.60	0.68	0.82	0.88
1	0.67	0.61	0.58	0.80	0.87

Robustness test regarding the labeling approach

Table 14 provides a sensitivity analysis regarding the consistency of the labeling approach when varying the number of clusters k.

Table 14: Labeling consensus given different values of k

This table shows the relative number of consistently labeled volatility interruptions when varying the number of clusters k. For the sake of clarity, we do not report the consensus in both ways as these are the same the other way round. Values are given as percentages.

	k=10	k=12	k=14	k=16	k=18
k=10	100				
k=12	71	100			
k=14	79	85	100		
k=16	86	77	85	100	
k=18	86	75	83	93	100

Robustness tests regarding the drivers of unnecessary interruptions

In this section, we provide different robustness tests for the results of the probit regression in Section 4. In Table 15, we present an OLS regression using stock-fixed effects in order to test for any stock-level effects. We use an OLS regression and not a probit model, as probit models tend to perform poorly with too many parameters, which is the case in this analysis. In Table 16, we provide the probit regressions using a binary variable indicating the turbulent COVID-19 market phases, in order to analyze the robustness of our results against such market-wide events. This binary variable is one for observations between February and May 2020. In Table 17, we present the results of the probit model using binary variables for each year in our sample period, in order to test for any time-dependent effects not currently captured by our model.

Table 15: OLS regression results with stock fixed effects

This table shows the results from OLS regressions analogous to the models described in Section 4. Here, the results are not based on probit regressions, but on OLS regressions, as the inclusion of stock fixed effects introduces 48 dummy variables, and probit regressions do not perform well when comprising a large amount of parameters. The relative spread and midpoint return volatility are given in basis points (bps). The level-1 depth and the trade volume is given in hundred thousand Euro. The message count is given in thousands. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	Market Quality	News and Contextual Factors	Full Model
const	0.4948***	0.2482***	0.4369***
rel_spread	−0.0047***		−0.0028***
leve_1_depth	0.0835***		0.0773***
trade_volume	−0.0003		−0.0005
midpoint_return_vola	−0.0511***		−0.0480***
message_count	−0.0083***		−0.0086***
news_count		−0.0002	−0.0001
news_relevance		−0.1479***	−0.0760**
news_relevance_interaction		0.0164	0.0260
unnecessary_interruptions		0.0043***	0.0030***
necessary_interruptions		−0.0023***	−0.0015***
fast_market_dummy		−0.0002***	−0.0001**
distance_to_static_barrier		0.2366***	0.0733*
trade_count_alt_venues		−0.0010***	0.0003
R-squared	0.0862	0.0569	0.0923
R-squared Adj.	0.0792	0.0492	0.0841
Stock FE	Yes	Yes	Yes
Accuracy	0.6392	0.6120	0.6428
Precision	0.6235	0.5922	0.6245
Recall	0.3449	0.2299	0.3644
Precision at Recall 20	0.6235	0.5922	0.6245

Table 16: Marginal effects of probit regression models including a COVID-19 dummy variable
This table shows the marginal effects of the proposed probit models. Marginal effects describe the absolute change in the output probability given a change in the independent variables. The COVID-19 dummy variable marks all observations between February and May of 2020. The relative spread and midpoint return volatility are given in basis points (bps). The level-1 depth and the trade volume are given in hundred thousand Euro. The message count is given in thousands. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

	Market Quality	News and Contextual Factors	Full Model
rel_spread	−0.0034***		−0.0027***
	(0.0009)		(0.0009)
leve_1_depth	0.0692***		0.0681***
	(0.0113)		(0.0112)
trade_volume	−0.0001		−0.0003
	(0.0003)		(0.0003)
midpoint_return_vola	−0.0614***		−0.0566***
	(0.0079)		(0.0081)
message_count	−0.0176***		−0.0175***
	(0.0023)		(0.0024)
covid_dummy	−0.0620***	−0.0360*	−0.0086
	(0.0157)	(0.0203)	(0.0203)
news_count		−0.0004	−0.0003
		(0.0002)	(0.0002)
news_relevance		−0.1887***	−0.1097***
		(0.0348)	(0.0342)
news_relevance_interaction		0.0353	0.0468
		(0.0309)	(0.0290)
unnecessary_interruptions		0.0044***	0.0033***
		(0.0010)	(0.0010)
necessary_interruptions		−0.0022***	−0.0016***
		(0.0004)	(0.0004)
fast_market_dummy		−0.0002***	−0.0001**
		(0.0001)	(0.0001)
distance_to_static_barrier		0.2269***	0.0404
		(0.0369)	(0.0379)
trade_count_alt_venues		−0.0014***	0.0004
		(0.0002)	(0.0003)
Pseudo R2	0.0645	0.0332	0.0689
Accuracy	0.6206	0.5945	0.6161
Precision	0.5484	0.4930	0.5373
Recall	0.3538	0.0757	0.3692
Precision at Recall 20	0.5484	0.5035	0.5373

Table 17: Marginal effects of probit regression models including dummy variables for each year

This table shows the marginal effects of the proposed probit models. Marginal effects describe the absolute change in the output probability given a change in the independent variables. The relative spread and midpoint return volatility are given in basis points (bps). The level-1 depth and the trade volume are given in hundred thousand Euro. The message count is given in thousands. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

	Market Quality	News and Contextual Factors	Full Model
rel_spread	−0.0034***		−0.0023**
	(0.0009)		(0.0009)
level_1_depth	0.0731***		0.0709***
	(0.0113)		(0.0113)
trade_volume	0.0001		−0.0002
	(0.0003)		(0.0003)
midpoint_return_vola	−0.0592***		−0.0570***
	(0.0079)		(0.0082)
message_count	−0.0198***		−0.0191***
	(0.0024)		(0.0025)
year_2020	−0.0693***	−0.0692**	−0.0387
	(0.0266)	(0.0283)	(0.0278)
year_2021	−0.0354	0.0069	−0.0443
	(0.0308)	(0.0314)	(0.0310)
year_2022	0.0200	0.0126	0.0092
	(0.0263)	(0.0266)	(0.0266)
year_2023	−0.0358	−0.0338	−0.0477*
	(0.0283)	(0.0287)	(0.0285)
year_2024	−0.0823***	−0.0656**	−0.0950***
	(0.0288)	(0.0294)	(0.0291)
news_count		−0.0004	−0.0003
		(0.0002)	(0.0002)
news_relevance		−0.1864***	−0.1047***
		(0.0348)	(0.0343)
news_relevance_interaction		0.0370	0.0472
		(0.0307)	(0.0290)
unnecessary_interruptions		0.0040***	0.0029***
		(0.0010)	(0.0010)
necessary_interruptions		−0.0021***	−0.0015***
		(0.0004)	(0.0004)
fast_market_dummy		−0.0002***	−0.0001**
		(0.0001)	(0.0001)
distance_to_static_barrier		0.2062***	0.0389
		(0.0366)	(0.0373)
trade_count_alt_venues		−0.0015***	0.0005*
		(0.0002)	(0.0003)
Pseudo R2	0.0679	0.0365	0.0727
Accuracy	0.6264	0.6098	0.6289
Precision	0.5585	0.5417	0.5586
Recall	0.3667	0.2329	0.3950
Precision at Recall 20	0.5585	0.5417	0.5586