Working Paper Series

No 152

Price dislocations: insights from trade repository data

by
Albert J. Menkveld
Ion Lucas Saru
Shihao Yu

ESRB
European Systemic Risk Board
European System of Financial Supervision

Abstract

This paper identifies price dislocation events in EuroSTOXX 50 futures, i.e., periods marked by high absolute returns. Combining public limit order book data with confidential trade repository data collected under the European Market Infrastructure Regulation (EMIR), we analyze market conditions around such dislocations. We find that price dislocations are accompanied by an increase in trading volume, and in the number of trades. EMIR data enables us to identify who participates in these trades, which allows us to tell if the volume increase is driven by fewer investors trading more, i.e., a more concentrated market, or by more investors participating. The latter could be argued to be a sign of a resilient market. We find evidence in support of such resilience, because the Herfindahl-Hirschman Index declines, both on the liquidity-demand and the liquidity-supply side. Our results further show that, contemporaneously, public order book variables explain most of the price dislocation events; adding private EMIR data contributes relatively little. We further find that *predicting* price dislocations is extremely hard, even after adding private EMIR data to public order book data.

Keywords: Price dislocations, EuroSTOXX 50 index futures, market concentration.

JEL codes: G14, G18.

Non-technical summary

This paper identifies price dislocation events in EuroSTOXX 50 futures, i.e., periods marked by high absolute returns. Combining public limit order book data with confidential trade repository data collected under the European Market Infrastructure Regulation (EMIR), we analyze market conditions around such dislocations.

We find that price dislocations are accompanied by an increase in trading volume, and in the number of trades. EMIR data enables us to identify who participates in these trades, which allows us to tell if the volume increase is driven by fewer investors trading more, i.e., a more concentrated market, or by more investors participating. The latter could be argued to be a sign of a resilient market. We find evidence in support of it; the market becomes less concentrated during price dislocations.

We further attempt to *predict* price dislocations with data sampled at a ten-second frequency. We find that it is extremely challenging to do so, even after including the (private) EMIR data.

Explaining price dislocations, however, proves more successful. We are able to explain up to 59% of price-dislocation returns. Public order book data explain the lion's share of it. The contribution of EMIR data is modest.

1 Introduction

Large price movements in financial markets can be driven by liquidity supply (for instance, liquidity providers withdrawing from markets) or liquidity demand (for instance, a trader liquidating a large position). In this paper, we identify such large price movements, or price dislocation events, and analyze trading during these events by supplementing limit order book (LOB) data with trade repository data collected under the European Market Infrastructure Regulation (EMIR). In contrast to commercially available LOB data that do not contain the identities of traders, EMIR data identifies the counterparty to each trade, both on- and off-exchange. As such, it allows us to study *disaggregated* trading patterns around price dislocation events and shed light on whether the dislocations are associated with, for example, the entry and exit of active traders in the market and the level of market concentration.

We propose a methodology to identify price dislocation events in financial markets, accounting for market volatility. Our approach to identifying price dislocations builds on the statistical process control (SPC) approach (Page, 1954). When applying SPC, we do account for conditional market volatility as estimated from a GARCH model. We consider a sudden large price change only a "price dislocation" if it is high relative to predicted volatility (i.e., conditional volatility). We apply our methodology to trading in an actively traded derivatives contract, EuroSTOXX 50 index futures, and identify 2637 price dislocation events in the period January 30, 2020, through December 29, 2023, with returns ranging from -96.18 to 99.43 basis points (bps).

Zooming in on price dislocation events reveals that, on average, they are accompanied by a decline in market concentration as well as an increase in the number of (active) counterparties. Surprisingly, the number of counterparties increases on *both* sides of the market. Using the history of trades during the trading day, we show that the increase in the number of counterparties is partly driven by the entry of new traders who did not trade before the price dislocation took place.

Using principal component analysis (PCA), we find that trade repository variables capture additional information above and beyond what can be learned from LOB data. The important question, however, is whether this additional information helps explain and predict price dislocation events. Contemporaneous regressions of returns on LOB variables and trade repository variables show that trade repository variables contain relatively little incremental information. Having said that, they do provide some additional insight. First, *on average*, the market becomes more competitive during these events in the sense that market concentration declines, both on the liquidity-demanding and the liquidity-supplying side. Second, *conditional on a dislocation event occurring*, the size of the dislocation decreases in how competitive the market becomes on the liquidity-supplying side. In other words, the largest dislocations coincide with a relatively high concentration on the supplying side. These are dislocations where few have stepped in to supply liquidity.

We further run logistic regressions to test whether price dislocation events can be predicted and examine the predictive power of different trade variables. Our results reveal statistically significant coefficients on the numbers of buyers and sellers in the market, controlling for LOB variables. The predictive power, however, is weak.

Related literature. Pasquariello (2014) documents widespread dislocations in global financial markets under stressful market conditions. He constructs a monthly price dislocation measure based on the violation of non-arbitrage conditions. In contrast, we provide a high-frequency analysis of price dislocations, including an ability to measure market competitiveness.

A closely related literature focuses on the participation of high-frequency traders during dislocation periods. Kirilenko et al. (2017) zoom in on high-frequency trading around the flash crash. Anand and Venkataraman (2016) find that market makers withdraw from the market during unfavorable market conditions. Cespa and Vives (2017) document that the number of mini flash crashes increased over the past decade. Brogaard et al. (2018) focus on high-frequency trading around extreme price movements and Bellia et al. (2024) focus on liquidity provision of designated market makers around extreme price movements. Their results indicate that high-frequency traders and designated market makers keep supplying liquidity around extreme price movements. We take a broader perspective and consider all counterparties around price dislocation events. Our results show that the number of counterparties on both sides of the market increases during price dislocation events.

Similar to us, Andersen et al. (2001) detect price dislocations in intraday data and analyze the distribution of volatility over the trading day. Andersen, Todorov, and Zhou (2023) propose a measure to detect non-arbitrage violation in high-frequency data. We propose a simple methodology building on Menkveld and Yueshen (2019) to identify price dislocation events and identify trading patterns around these events.

The remainder of the paper is structured as follows. Section 2 discusses our methodology to identify price dislocation in financial markets. Section 3 presents our data. Empirical results are presented in Section 4. Finally, Section 5 concludes.

2 Methodology

Central to our analysis is the identification of price dislocation events. We identify price dislocations based on midquote returns at at 10-second frequency. Our measure builds on the statistical process control (SPC) approach (Page, 1954). This method has been applied, for instance, by Menkveld and Yueshen (2019). Consider the return process r_t, which has a mean of $\bar{r}$ and some variance

σ under normal condition, i.e., when it is under control. We recursively construct the upper cumulative sum as

$$s_{u,t} = \max\{0, s_{u,t-1} + r_t - (\bar{r} + k)\} \tag{1}$$

and similarly the lower cumulative sum as

$$s_{l,t} = \max\{0, s_{l,t-1} - r_t + (\bar{r} - k)\}, \tag{2}$$

where k is the slackness of the return process. In addition, we choose the control threshold h such that the process is out of control if $s_{\cdot,t} > h$. We follow Menkveld and Yueshen (2019) as well as Montgomery (2019) and set $h = 5\sigma$ and vary the slackness $k \in \{3\sigma, 4\sigma, 5\sigma, 6\sigma\}$.[1] We refer to instances when the return process is out of control as price dislocation events. Occurrences when $s_{l,t} > h$ correspond to negative return events and we refer to these events as negative price dislocations. Occurrences when $s_{u,t} > h$ correspond to positive return events and we refer to these events as positive price dislocations.

For instances that the return process is out of control for two or more consecutive time intervals, we keep only the first occurrence when we detect a price dislocation. We do not consider price dislocations that occur during the first 15 minutes of the trading day, as quotes have been documented to be unreliable during this period (Bogousslavsky, 2021).

We estimate σ based on a t-distributed GARCH(1,1) model. We estimate the model based on a rolling window and compute the daily one-step ahead conditional volatility forecasts and scale them to our sample frequency of 10-seconds. Using volatility forecasts based on daily returns in combination with a GARCH model alleviates concerns of computing (intraday) realized volatilities around price dislocation events or local non-arbitrage violations (Andersen, Todorov, and Zhou, 2023). We compute $\bar{r}$ based on the average 10-second return on the previous trading day, which is close to zero over our sample period.

Intuitively, we compute conditional percentiles of the return distribution. In high volatility periods, we expect prices and therefore returns to move more before we consider them to be "out of control". At the same time, we expect less movement in prices in low-volatility periods, such than already smaller price movements can constitute price dislocation events. In comparison to our approach, Brogaard et al. (2018) suggest identifying extreme price movements as the absolute returns above the 99.9th percentile of the absolute return distribution. As they note, their methodology identifies relatively more price dislocations in high-volatility periods versus low-volatility periods. In comparison, our price dislocations are more equally spread over high- and low-volatility periods.

3 Data

Our main analysis focuses on trading in EuroSTOXX 50 futures and combines data from two sources: limit order book (LOB) data from BMLL and trade

[1] This is at the restrictive end of the range recommended by Montgomery (2019).

repository data from the European Market Infrastructure Regulation (EMIR). Using LOB data, we are able to complement the transactions-level data from EMIR with an on-exchange reference price series and order book depth information. This follows the approach of, for example, Pinter (2022).

3.1 EMIR data

The EMIR data contain trade records for transactions involving European counterparties. Trade records are timestamped to the second. These data give a unique insight about how many market participants are active in the market at the same time. Level 3 order book information allows traders to infer, for instance, the number of orders at the best bid and at the best ask, as well as deeper in the order book. However, it does not contain information on the actual number of traders in the market or the type of traders. We clean the trade repository data from EMIR according to the procedure described in Appendix A.

Moreover, the data contain information about the quantities and notional amounts traded by the market participants, allowing us to compute the following variables

- Market Concentration: Herfindahl-Hirschman Index (HHI) based on notional amounts traded by all counterparties in the market.
- Buyer Concentration: HHI based on notional amounts traded by buying counterparties in the market.
- Seller Concentration: HHI based on notional amounts traded by selling counterparties in the market.
- Number Traders: Number of all counterparties in the market.
- Number Buyers: Number of buying counterparties in the market.
- Number Sellers: Number of selling counterparties in the market.
- Number Traders: Number of trades by all counterparties.
- Volume: Volume traded by all counterparties.

Menkveld and Saru (2024) show that clients appear relatively more informed than intermediaries at lower intraday frequencies. Moreover, buying or selling pressure from clients may cause intermediaries to adjust their prices in response to their inventory management, resulting in price pressures (Hendershott and Menkveld, 2014). Based on trader identities in EMIR as well as the list of exchange members published on Eurex' website,[2] we compute

- Client Buy Share: Share of client buy orders in all client orders.
- Cumulative Client Buy Share: Cumulative share of client buy orders in all client orders since the beginning of the trading day.

[2]See https://www.eurex.com/ex-en/trade/participant-lists/exchange-participants.

3.2 Order book data

We obtain Level-3 order book data from BMLL. These data serve two purposes: Firstly, we construct an on-exchange reference price series based on the prevailing midquote price. Secondly, we construct several order book variables to analyze market movements around price dislocation events. For the construction of order book variables, we follow Capponi and Yu (2024) and consider order book action variables (Brogaard, Hendershott, and Riordan, 2019) as well as limit order book state variables.

For the limit order book action variables, we follow the approach of Capponi and Yu (2024) and keep the quantity information of the order book events. Based on this, we construct the following variables:

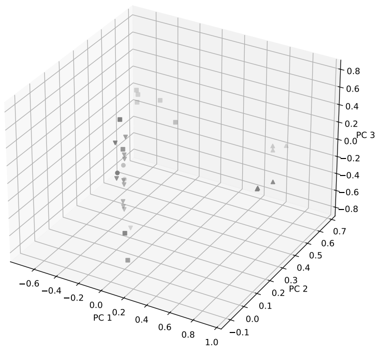
- BBO-Moving Trades: Market(able) buy or sell orders that result in trades moving BBO prices.
- Non-BBO-Moving Trades: Market(able) buy or sell orders that result in trades but do not move BBO prices.
- BBO-Improving Limit: Limit orders that improve upon the prevailing best bid or best ask price.
- BBO-Worsening Cancel: Cancel orders that worsen the prevailing best bid or best ask price.
- BBO-Depth Add Limit: Limit orders adding depth at the prevailing best bid or best ask price.
- BBO-Depth Remove Cancel: Cancel orders removing depth at the prevailing best bid or best ask price.
- Non-BBO-Depth Add Limit (5 prices): Limit orders adding depth outside the BBO but within the best five prices on the bid or ask side.
- Non-BBO-Depth Remove Cancel (5 prices): Cancel orders removing depth outside the BBO but within the best five prices on the bid or ask side.
- Non-BBO-Depth Add Limit: Limit orders adding depth outside of the best five prices.
- Non-BBO-Depth Remove Cancel: Cancel orders removing depth outside of the best five prices.

For the limit order book state variables, we compute

- BBO Imbalance: Difference in depth at the best bid and best offer.
- Non-BBO Imbalance (5 prices): Difference in cumulative depth at the five best bids and five best offers.
- Non-BBO Imbalance: Difference in cumulative depth at the bid and offer side beyond the five best prices.

Figure 1: Principal component analysis of variables

This figure plots the factor loadings of the variables derived from the order book data and the trade repository data from EMIR on their first three principal components. The principal components are computed based on a window around price dislocation events, starting 5 minutes before and ending 5 minutes after price dislocation events occur. We use a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. Order book state variables are denoted by circles, other order book variables are denoted by downward pointing triangles, concentration measures from EMIR are denoted by upward pointing triangles, and other variables derived from EMIR are denoted by squares.



Finally, such as Capponi and Yu (2024), we compute

- BBO Queue Length Imbalance: Difference in the number of orders at the best bid and the number of orders at the best offer.

Given the different nature of our variables derived from the order book data and the trade repository data from EMIR, the question arises to which extent they capture the same information. To shed light on this question, we take the first three principal components of the variables in the 10-minute windows around price dislocation events. The first three principal components explain 41% of the variation in the different variables. We exhibit factor loadings of the different variables in Figure 1. As can be seen, the variables form different clusters, with the concentration measures from trade repository data as well as the variables capturing the number of counterparties, volume, and the number of trades forming distinct clusters with respect to the remaining variables.[3]

[3] Figure 6 in Appendix B shows consistent results when we apply an exploratory factor analysis.

3.3 Sample construction

Our sample period spans from February 2019 to December 2023. As the implementation of our SPC measure requires the estimation of a GARCH model (see Section 2), we apply our SPC measure to the period January 30, 2020 to December 29, 2023. We select the closest-to-maturity contract and roll over to the next contract 5 days before expiration of the contract. This follows Andersen et al. (2007). Our results are robust to contract selection based on the highest-volume contract.

We implement our analysis at the 10-second frequency. With this, we follow the same trade-off as described in Brogaard et al. (2018). We want to sample at a sufficiently high frequency to identify dislocations with a short duration. At the same time, sampling at a too high frequency may split price changes between several time intervals, making it more difficult to detect them. Note, however, the statistical nature of our approach to identifying price dislocation events. As the return process is subject to some slackness, our methodology also captures dislocations that are split between multiple time intervals as long as the individual interval returns are not too small in absolute value.

4 Results

In this section we first present summary statistics on the price dislocation events we identify. We then show results on market dynamics around price dislocation events, before turning to the question whether price dislocation events are predictable. Finally, we construct a fragility index based on large language models (LLMs).

4.1 Price dislocation events

Figure 2 plots the price dislocation events we identify for a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$ over the trading day as well as over our sample period. To give context, we also plot conditional volatility forecasts from a GARCH model as described in Section 2.

Panel 2a shows that we identify most price dislocation events around the middle of the trading day. Price dislocations do not cluster (both in frequency or in magnitude) around the market open or close.[4]

Over time, we identify price dislocation events in both high-volatility as well as low-volatility periods (see Figure 2b). While price dislocation events do not cluster in high-volatility periods, the magnitude of price dislocations

[4]We also consider potential patterns around the opening of US markets. Out of the total number of 2637 price dislocations that we identify, 822 occur after US markets open and 116 occur in the first 15 minutes after US markets open. The average return of positive price dislocations after US markets open is 11.44bps (overall sample: 12.29bps, first 15 minutes after US markets open: 10.09bps). The average return of negative price dislocations after US markets open is −12.04bps (overall sample: −12.40bps, first 15 minutes after US markets open: −9.87bps).

increases as conditional volatility forecasts increase. This reflects the intuition that we expect prices to move relatively more in high-volatility periods than in low-volatility periods, regardless of whether they dislocate or not. Therefore, 10-second returns have to be larger in magnitude for them to constitute a price dislocation event in high volatility periods.

Figure 7 in Appendix B shows the distribution of price dislocations as well as absolute returns above the 99.9th percentile of the absolute return distribution. While the number of identified events is comparable, the distributions differ. Identifying price dislocations based on the absolute return distribution over-samples high-volatility periods relative to our method.

4.2 Market dynamics around price dislocations

First, we analyze market dynamics around price dislocation events. In particular, we are interested whether there are any distinct patterns in order book or trade repository variables. Therefore, we first construct a placebo sample.

Placebo sample

We construct a placebo sample to identify whether the patterns we identify are specific to the time of the day, following the approach of van Kervel and Menkveld (2019). Therefore, we match four trade variables from the market open until 5 minutes before the price dislocation event takes place. During the matching procedure, we make sure to match with days on which we do not identify any price dislocation events, accounting for the fact that price dislocation-days may be inherently different. We match on the following trade variables: return, volume, order imbalance, and realized volatility (based on 1-minute returns).

Following van Kervel and Menkveld (2019) we match price dislocation events based on the nearest-neighbor method, measuring the distance in standard deviations rather than in percentages.

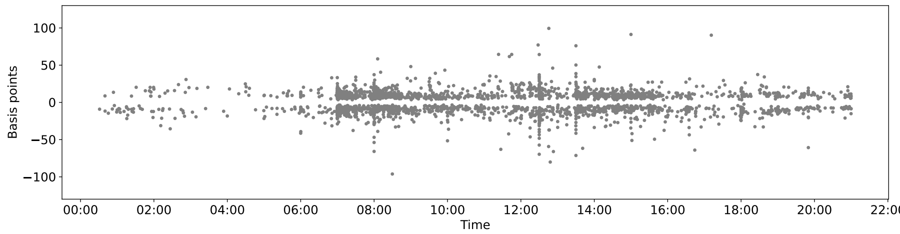
Our matching procedure addresses the nature of the price dislocation events we identify. If the price dislocation events we capture are a systematic event that occurs regularly around the same time of the trading day, we expect similar patterns in both the placebo sample and the dislocation sample. Similarly, if the occurrence of price dislocations is fully determined by the trading patterns since market opening (as captured by our matching variables), we expect similar patterns in our treatment and placebo samples. However, if price dislocation events are local and/or idiosyncratic events, we expect different patterns in the treatment and placebo samples in a window around the price dislocation events.

Market dynamics

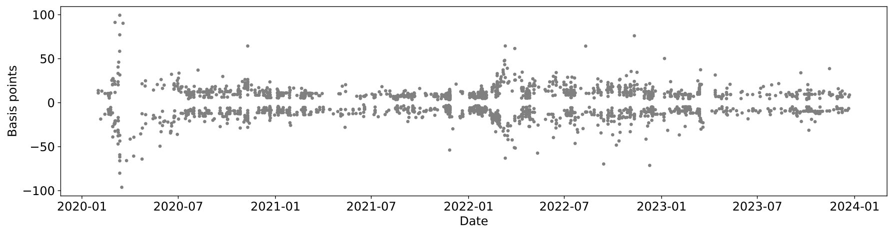
Having established the placebo sample as the reference group, we present market dynamics in the vicinity of positive price dislocations in Figure 3 and market dynamics in the vicinity of negative price dislocation in Figure 4. We consider a

Figure 2: Price dislocation events and conditional volatility forecasts

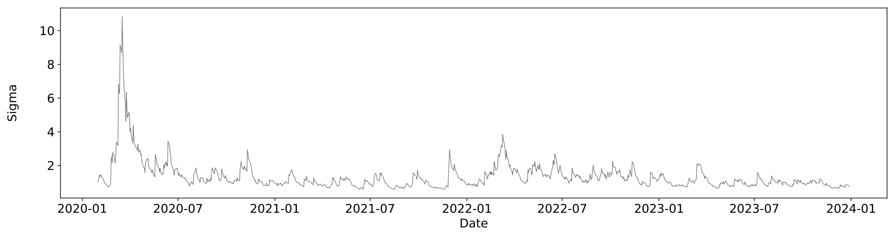
Panel (a) plots price dislocation events for a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$ over the trading day. For each dislocation event, we plot the corresponding 10-second return in basis points. Panel (b) plots price dislocation events for the same slackness parameter and control threshold over the sample period. Again, we plot the corresponding 10-second return in basis points. Panel (c) shows one-step ahead conditional volatility forecasts from a GARCH model.



(a) Price dislocations during day



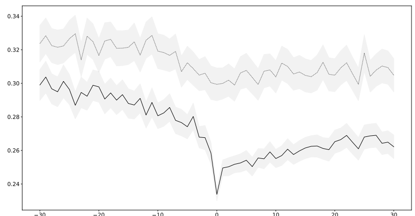
(b) Price dislocations over time



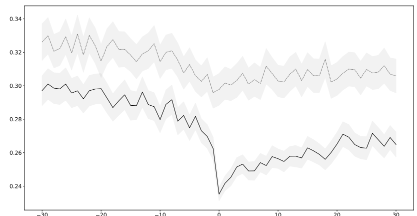
(c) Conditional volatility forecasts over time

Figure 3: Market dynamics in the vicinity of positive price dislocation events

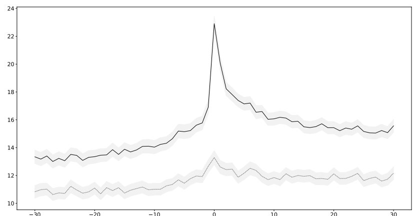
The figure plots different market variables for a 10-minute interval around positive price dislocation events with a slackness parameter of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. Solid lines refer to the treatment sample and dotted lines to the placebo sample, with shaded areas being the corresponding 95% confidence intervals. We plot the variables for each 10-second interval, starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.



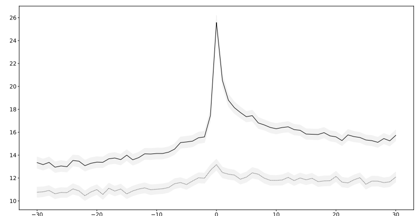
(a) Buyer Concentration



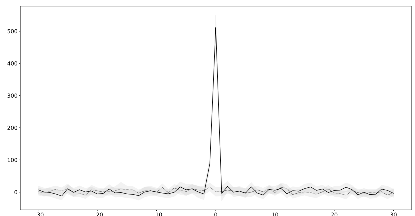
(b) Seller Concentration



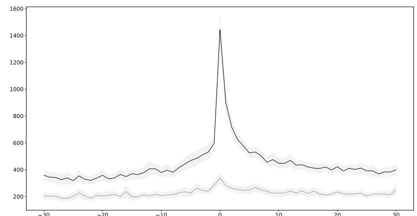
(c) Number Buyers



(d) Number Sellers



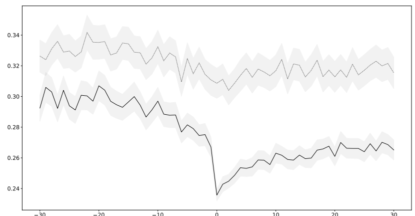
(e) Order Imbalance



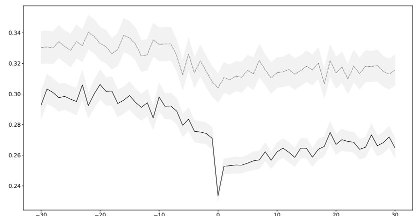
(f) Volume

Figure 4: Market dynamics in the vicinity of negative price dislocation events

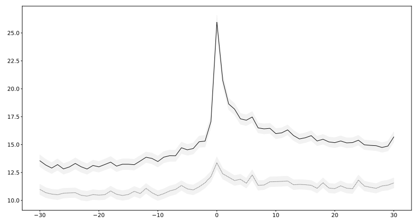
The figure plots different market variables for 10-minute interval around negative price dislocation events with a slackness parameter of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. Solid lines refer to the treatment sample and dotted lines to the placebo sample, with shaded areas being the corresponding 95% confidence intervals. We plot the variables for each 10-second interval, starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.



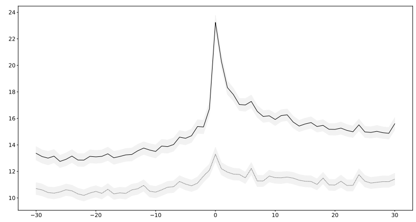
(a) Buyer Concentration



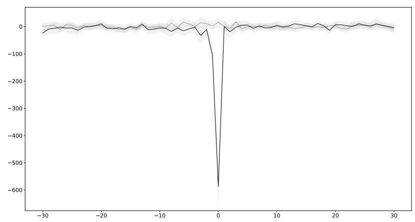
(b) Seller Concentration



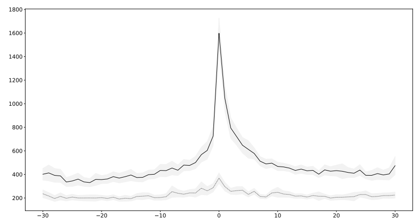
(c) Number Buyers



(d) Number Sellers



(e) Order Imbalance



(f) Volume

window starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place. We compute all market variables for each 10-second interval in this window. In the following, we distinguish between positive and negative price dislocations both to identify whether there are different patterns for both types of price dislocation events and well as to account for the directionality of trade variables, such as order imbalances.[5]

The variables we consider are buyer concentration, seller concentration, the number of buyers, the number of sellers (all derived from trade repository data), order imbalance, and trading volume (both derived from order book data).[6] Overall, the patterns are comparable for both positive and negative price dislocation events.

It is notable that for both positive and negative price dislocation events, buyer concentration as well as seller concentration decrease in the vicinity price dislocation events. At the same time, both the number of buying counterparties as well as the number of selling counterparties increases. Comparing the treatment sample with the placebo sample reveals that buyer concentration and seller concentration are on a lower level during the entire 10-minute interval around price dislocation events. Similarly, the number of buyers and the number of sellers is elevated over the entire 10-minute interval on price dislocation days.

Comparing the pre-dislocation period with the post-dislocation period reveals several patterns. On average, both buyer concentration and seller concentration remain on a lower level in the post-dislocation period, both relative to the pre-dislocation period as well as the placebo sample. Likewise, the number of buying counterparties and the number of selling counterparties remains elevated in the post-dislocation period.

For all trade repository variables, we observe a gradual pattern running up to the price dislocation event. For both positive and negative price dislocations, market concentrations on both the buyer and seller sides start to decrease approximately 100 seconds before the price dislocation events occurs. The number of buyers starts to increase for positive price dislocation events, and the number of sellers starts to increase for negative price dislocation events. There is also an increase in the number of sellers (buyers) for positive (negative) price dislocation events, however, this increase is less pronounced and occurs on average approximately 50 seconds before price dislocation occurs.

Order imbalance exhibits a pronounced pattern in close proximity to the price dislocation event. Positive price dislocation events are accompanied by a positive order imbalance, while negative price dislocation events are accompanied by a negative order imbalance. While the trade repository variables show a pre-dislocation and post-dislocation trend, the pattern is less pronounced for order imbalances. For positive price dislocation events, order imbalance turns

[5]Note that the results in this Section are average results for all price dislocation events in our sample period. In Appendix C, we investigate whether the patterns were comparable in March 2020, as this period was subject to margin call events in European equity derivatives markets (ESRB, 2020).

[6]The pattern for trading volumes is comparable when using data from EMIR.

significantly positive relative to the placebo sample in the 10-second interval preceding the price dislocations. Similarly, order imbalance turns significantly negative in the 10-second interval preceding negative price dislocation events. We do not find significant differences in order imbalances for all other 10-second intervals in the 10-minute window around both positive and negative price dislocation events.

Trading volumes spike during the 10-second intervals corresponding to the price dislocation events. There is a significant increase in trading volumes starting approximately 50 seconds before price dislocations occur. Similarly, trading volumes remain elevated in the post-dislocation period for 60 seconds before returning to their pre-dislocation levels. As for the trade repository variables, there is a significant difference in volumes between the treatment and placebo samples over the entire 10-minute interval around price dislocation events.

Our finding that both the number of buying and selling counterparties increases during all price dislocation events complements the findings of Brogaard et al. (2018). They document that high-frequency traders provide liquidity around extreme price movements rather than exiting the market. Our results contribute to the literature by showing that the overall number of sellers (buyers) increases during positive (negative) price dislocation events. At the same time, market concentration decreases on both sides of the market. Therefore, also the effective number of sellers (buyers) increases. Furthermore, this is aligned with Biais, Declerck, and Moinas (2016) who show that proprietary traders provide liquidity during stressed market conditions.

The finding that there are significant differences between the dislocation sample and the placebo sample during the 10-minute interval around price dislocation events for all trade repository variables as well as trading volume suggests that price dislocation days are different from non-price dislocation days. Moreover, this raises the possibility that there is some predictability around price dislocation events. We explore this possibility in more detail in Section 4.3.

Permanent and transitory dislocations

Price dislocation events may differ in their nature. Events triggered by news events may result in permanent price movements, while events occurring for liquidity reasons may result in transitory price movements. Therefore, we distinguish between permanent and transitory price dislocation events. Market participants may be more inclined to provide liquidity during transitory price dislocations than during permanent price dislocations.

We identify permanent and transitory price dislocations following Brogaard et al. (2018). We classify price dislocation events that revert by more than 2/3 at the end of a 30-minute period as transitory price dislocation events. Conversely, price dislocation events that revert by less than 1/3 after 30 minutes are classified as permanent price dislocations. We drop price dislocation events that are neither classified as transitory nor as permanent. For a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$ this amounts to 6.07% of all price dislocation events.

Figures 8 and 9 in Appendix B present results for the number of sellers during positive and buyers during negative price dislocation events, respectively. We find that during both transitory and permanent price dislocation events the number of both buying and selling counterparties increases. This is not aligned with the hypothesis that liquidity providers enter the market around transitory price movements but not around permanent price movements. Rather, both types of price dislocations are accompanied by an increase in the number of market participants on the liquidity supplying side.

Return patterns

Next, we turn to analyzing return patterns in the vicinity of price dislocation events. Again, we consider a 10-minute window around price dislocation events. Moreover, this analysis addresses the question whether there is information in trade repository data beyond the information contained in LOB variables.[7] As discussed in Section 3, the trade repository data contain information on the number of counterparties in the market over time. Given that the risk-bearing capacity of individual counterparties is limited, this data may be informative beyond order book information.[8]

Therefore, we perform contemporaneous regressions of returns on our set of order book and trade repository variables. Results for positive price dislocations are presented in Table 1 and results for negative price dislocations are presented in Table 2. In all regression specifications, all explanatory variables are standardized to have a mean of zero and a standard deviation of one to facilitate economic interpretability. In addition, we vary the slackness from $k = 3\sigma$ to $k = 6\sigma$ to investigate robustness as the magnitude of the dislocations increases.

All results reveal that the trade repository variables have explanatory power with respect to contemporaneous returns, both individually as well as controlling for liquidity in the order book. For both positive and negative price dislocation events, higher buyer concentration is associated with more negative returns and higher seller concentration is associated with more positive returns. Both findings are intuitive: as buyer (seller) concentration increases, the effective number of buyers (sellers) decreases. Therefore, the market imbalance is split among less buyers (sellers). This is consistent with an inventory management channel such as Hendershott and Menkveld (2014).

Across all specifications, an increase in the number of buyers (sellers) is associated with an decrease (increase) in conditional expected returns. This finding suggests that periods with less buyers (sellers) are associated with more positive (negative) returns. Price dislocations may be idiosyncratic events that are less predictable for liquidity providers (Bessembinder et al., 2016) or events that are more likely driven by a few large buyers (sellers) (Menkveld and Yueshen, 2019).

[7]Note that controlling for different order book measures controls for different dimensions of liquidity in the market.

[8]Moreover, as we show in Figure 1, the concentration measures derived from EMIR form a distinctive cluster of variables.

Table 1: Contemporaneous regressions for positive price dislocation events

This table presents results for contemporaneous regressions of returns on LOB variables and EMIR variables for a window ranging from 5 minutes before to 5 minutes after a price dislocation takes place. We vary the slackness parameter by setting $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. Standard errors are robust to heteroscedasticity as well as autocorrelation up to 61 lags and are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.253***	0.253***	0.253***	0.253***	0.253***	0.253***
	(0.019)	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
Market Concentration	−0.063	0.092***				0.001
	(0.055)	(0.028)				(0.031)
Buyer Concentration	−0.419***	−0.152***				−0.137***
	(0.044)	(0.020)				(0.026)
Seller Concentration	0.504***	0.031				0.151***
	(0.040)	(0.021)				(0.021)
Number Traders	−0.201		−0.279			−0.140
	(0.270)		(0.189)			(0.197)
Number Buyers	−1.959***		−0.202			−0.365**
	(0.203)		(0.129)			(0.142)
Number Sellers	2.104***		0.593***			0.429***
	(0.159)		(0.096)			(0.103)
Client Buy Share	1.165***			0.182***		0.161***
	(0.031)			(0.033)		(0.033)
Cum. Client Buy Share	−0.087***			−0.030**		−0.026*
	(0.019)			(0.013)		(0.013)
Volume	0.005				−0.123***	−0.111***
	(0.054)				(0.039)	(0.036)
Number Trades	0.702***				0.291***	0.353***
	(0.127)				(0.068)	(0.093)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
R^2 (%)	15.26	57.22	57.35	57.32	57.45	57.75
N	61670	61670	61670	61670	61670	61670
	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.370***	0.370***	0.370***	0.370***	0.370***	0.370***
	(0.057)	(0.039)	(0.039)	(0.039)	(0.038)	(0.037)
Market Concentration	0.063	0.206***				0.089
	(0.148)	(0.075)				(0.090)
Buyer Concentration	−0.602***	−0.203***				−0.213***
	(0.130)	(0.052)				(0.079)
Seller Concentration	0.533***	−0.036				0.133**
	(0.099)	(0.061)				(0.054)
Number Traders	0.447		0.024			0.316
	(0.916)		(0.527)			(0.560)
Number Buyers	−3.831***		−0.668*			−1.035***
	(0.615)		(0.351)			(0.387)
Number Sellers	3.083***		0.716***			0.526*
	(0.636)		(0.265)			(0.308)
Client Buy Share	1.184***			0.006		−0.037
	(0.074)			(0.057)		0.062
Cum. Client Buy Share	−0.112**			0.005		0.001
	(0.047)			(0.034)		(0.033)
Volume	0.282				−0.371**	−0.324**
	(0.193)				(0.173)	(0.154)
Number Trades	1.332***				0.521**	0.640**
	(0.344)				(0.233)	(0.284)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
R^2 (%)	15.46	58.73	58.86	58.71	59.04	59.28
N	13609	13609	13609	13609	13609	13609

Table 2: Contemporaneous regressions for negative price dislocation events

This table presents results for contemporaneous regressions of returns on LOB variables and EMIR variables for a window ranging from 5 minutes before to 5 minutes after a price dislocation takes place. We set the slackness parameter $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. Standard errors are robust to heteroscedasticity as well as autocorrelation up to 61 lags and are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.307^{***}	-0.307^{***}	-0.307^{***}	-0.307^{***}	-0.307^{***}	-0.307^{***}
	(0.019)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Market Concentration	0.057	-0.028				0.063^{*}
	(0.058)	0.030				(0.037)
Buyer Concentration	-0.593^{***}	-0.079^{***}				-0.251^{***}
	(0.045)	(0.024)				(0.035)
Seller Concentration	0.492^{***}	0.156^{***}				0.170^{***}
	(0.041)	(0.022)				(0.024)
Number Traders	-0.067		0.199			0.144
	(0.265)		(0.214)			(0.215)
Number Buyers	-2.340^{***}		-0.841^{***}			-0.732^{***}
	(0.214)		(0.158)			(0.169)
Number Sellers	2.337^{***}		0.477^{***}			0.550^{***}
	(0.169)		(0.122)			(0.130)
Client Buy Share	1.130^{***}			0.320^{***}		0.276^{***}
	(0.028)			(0.037)		(0.036)
Cum. Client Buy Share	-0.052^{**}			-0.024^{*}		-0.011
	(0.021)			(0.013)		(0.015)
Volume	-0.019				0.087^{***}	0.080^{***}
	(0.032)				(0.031)	(0.028)
Number Trades	-0.544^{***}				-0.224^{***}	-0.231^{**}
	(0.170)				(0.076)	(0.115)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
R^2 (%)	14.44	51.04	51.37	51.40	51.19	51.88
N	63455	63455	63455	63455	63455	63455

	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.333^{***}	-0.333^{***}	-0.333^{***}	-0.333^{***}	-0.333^{***}	-0.333^{***}
	(0.057)	(0.038)	(0.040)	(0.038)	(0.040)	(0.040)
Market Concentration	0.140	0.053				0.140
	(0.150)	(0.079)				(0.091)
Buyer Concentration	-0.772^{***}	-0.125^{**}				-0.318^{***}
	(0.106)	(0.056)				(0.090)
Seller Concentration	0.532^{***}	0.088				0.082
	(0.114)	(0.056)				(0.060)
Number Traders	-1.240		-0.939			-0.971
	(0.934)		(0.612)			(0.657)
Number Buyers	-2.704^{***}		-0.222			-0.205
	(0.696)		(0.442)			(0.494)
Number Sellers	3.911^{***}		0.971^{***}			1.065^{***}
	(0.524)		(0.337)			(0.378)
Client Buy Share	1.231^{***}			0.237^{***}		0.199^{***}
	(0.064)			(0.073)		(0.072)
Cum. Client Buy Share	-0.044			-0.026		-0.008
	(0.045)			(0.034)		(0.035)
Volume	-0.368				0.052	0.115
	(0.285)				(0.206)	(0.205)
Number Trades	-0.497				-0.263	-0.272
	(0.583)				(0.297)	(0.369)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
R^2 (%)	12.85	52.06	52.29	52.18	52.18	52.52
N	14675	14675	14675	14675	14675	14675

Contemporaneous returns are positively associated with the client buy share (i.e., the share of client buy orders in all client orders). In addition, as we show in Figure 10 in Appendix B, the client buy share increases during positive price dislocation events and decreases during negative price dislocation events. Together, these results suggest that client orders trade in same direction of price movements around price dislocation events, thus demanding liquidity.

When varying the slackness from $k = 3\sigma$ to $k = 6\sigma$, some variables lose traction relative to the specification with $k = 3\sigma$. Buyer concentration as well as the number of buyers (sellers) appear to be the most important drivers of returns among trade repository variables in the vicinity of positive (negative) price dislocation events.

Comparing the results controlling for order book variables with the results without order book variables reveals two findings. First, public order book variables explain most of the price dislocation events; adding trade repository data adds relatively little. Second, there is incremental information in the trade repository variables above and beyond what is captured by the liquidity metrics from the order book variables.

We estimate logistic regressions of an indicator variable for price dislocation events on the different order book and trade repository variables. Results by direction of the price dislocation events are reported in Tables 8 and 9 in Appendix B. The results are mostly aligned with the results from contemporaneous regressions of returns on order book and trade repository variables.

Are counterparties entering the market?

The contemporaneous return regressions presented in Tables 1 and 2 indicate that an increase in the number of buyers (sellers) is associated with a decrease (increase) in contemporaneous returns. In addition, the patterns in Figures 3 and 4 show an increase in the number of buyers and sellers in the vicinity of price dislocation events.[9] This raises the question whether the buying and selling counterparties we observe in the market during price dislocation events were present in the market before the price dislocations occur, or whether they entered the market for the first time during the price dislocation events.

We compute the number of counterparties that trade for the first time on a trading day during the price dislocation event (first-time traders), the number of counterparties that act as buyers for the first time during the trading day (first-time buyers), and the number of counterparties that act as sellers for the first time during the trading day (first-time sellers). Again, we distinguish between positive and negative price dislocation events.

We present our results graphically in Figures 11 and 12 in Appendix D. The results are comparable for positive and negative price dislocation events. On average, the number of first-time traders, first-time buyers, and first-time sellers

[9] Additionally, this is consistent with the logistic regression results presented in Tables 8 and 9.

increases in the 10-second interval when the price dislocation event occurs.[10] The results indicate that traders enter the market on both sides during price dislocation events. Therefore, during negative price dislocation events, traders enter the market not only to sell, but also to buy; thus providing liquidity.

On-exchange and off-exchange trades

The trade records in EMIR trade repository data pertain to both on-exchange as well as off-exchange trades. As a result of trading relationships and relationship discounts (Czech et al., 2021; Jurkatis et al., 2022), off-exchange trading volumes may increase during volatile periods. In particular, we test the null hypothesis that there is no increase in off-exchange trading around price dislocation events. As there is no identifier for on-exchange and off-exchange trades in EMIR, we estimate the regression

$$Volume_t^{EMIR} = \beta_0 + \beta_1 Volume_t^{Eurex} + \beta_2 \mathbb{I}(spc)_t + \beta_3 \mathbb{I}(spc)_t Volume_t^{Eurex} + \varepsilon_t.$$

β_1 is the volume multiplier. If there is no off-exchange trading, we expect $\beta_1 \leq 2$ (as both sides of the trade are reported).[11] The coefficient β_3 captures whether there is any change in the volume multiplier during price dislocation events. We consider two scenarios: only the price dislocation events themselves as well as the 10 minutes around price dislocation events. For both the price dislocation events themselves as well as a 10-minute window around price dislocation events, we reject the null hypothesis $\beta_1 \leq 2$ but cannot reject the null hypothesis $\beta_3 = 0$. That is, we do not find evidence for a change in off-exchange trading during price dislocation events.

4.3 Are price dislocations predictable?

As discussed in the previous Section, we observe distinctive patterns in various variables in the run-up to price dislocation events. These findings raise the question whether price dislocation events are predictable.

We investigate this question by estimating in-sample predictive logistic regressions

$$Prob(SPC = 1)_t = \alpha + \boldsymbol{X}_{t-1}^{EMIR} \boldsymbol{\beta}^{EMIR} + \boldsymbol{X}_{t-1}^{LOB} \boldsymbol{\beta}^{LOB} + \beta^{Ret} Ret_{t-1} + \varepsilon_t \quad (3)$$

for an interval starting 5-minutes before a price dislocation occurs and ending with the 10-second interval during which the price dislocation takes place. We include different sets of one-period lagged variables: the predictor variables

[10] For first-time traders, first-time buyers, and first time sellers we test the null hypothesis that the respective variable is zero in the 10-second interval when the price dislocation event takes place. For both positive and negative price dislocation events as well as for slackness parameters $k = 3\sigma$ and $k = 6\sigma$, we reject the null hypothesis for each variable at all conventional significance levels.

[11] However, note that we only observe trade records involving European counterparties in EMIR. Therefore, if a European counterparty trades with a non-European counterparty, we observe only one side of the trade.

derived from the trade repository data, the predictor variables derived from the LOB data, and the one-period lagged 10-second return. With this, we address the question whether price dislocation events are predictable in-sample in the run-up before they occur.

Results for positive price dislocation events are presented in Table 3 and results for negative price dislocation events are presented in Table 4. Overall, evidence for predictability of price dislocation events is weak. For positive price dislocation events and a slackness of $k = 3\sigma$, an increase in the number of buyers and sellers as well as in increase in the client buy share is associated with an increase in the probability of the occurrence of a price dislocation event. These findings are consistent with the patterns described in Section 4.2. For relatively larger positive price dislocations with a slackness of $k = 6\sigma$, we do not find evidence that the trade repository variables have predictive power.

The findings for positive and negative price dislocations are largely comparable for a slackness of $k = 3\sigma$. For negative price dislocations with a slackness of $k = 6\sigma$, we find evidence consistent with weak predictive ability of the number of sellers as well as the number of trades. The finding that a higher (lower) client buy share is associated with a higher probability of a positive (negative) price dislocation event in the next period is additional evidence consistent with liquidity demand by clients around price dislocation events.

4.4 Predicting crash risk from large language models

In addition to the EMIR database and on-exchange LOB data, textual datasets such as financial and macroeconomic news can also provide useful information for predicting stock market crash risk. Large language models (LLMs) such as ChatGPT are trained on a vast volume of textual datasets and have a superior ability to extract such information from financial and macroeconomic news. In the following exercise, we first collect financial and macroeconomic news pertinent to the ESX and DAX indices and use ChatGPT to interpret the news and construct a fragility or crash risk index for each stock index.

Scraping news from news outlet

We scrape news articles from the website of the Financial Times for ChatGPT to evaluate. Specifically, for a particular index, we search all news articles containing the relevant keywords. For example, when collecting news articles for DAX, a German stock market index, we search keywords such as "German economy", "German trade", "German inflation", "German unemployment", "German stocks", and "German markets". Instead, when collecting news articles for ESX, a stock index of Eurozone stocks, we search keywords such as "European economy", "European trade", "European inflation", "European unemployment", "European stocks", and "European markets".

For each news article, we obtain its timestamp, headline, and standfirst. Below is an example news article about the German economy on July 1, 2024:

Table 3: Predictive regression results for positive price dislocation events

This table presents results from predictive logistic regressions for positive price dislocation events on one-period lagged LOB variables, EMIR variables, and one-period lagged returns. We consider a window starting 5 minutes before the price dislocation event occurs and ending with the price dislocation event. We vary the slackness parameter by setting $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. *Controls* indicate that the specification controls for both the LOB variables as well as the one-period lagged return. Robust standard errors are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-3.358^{***}	-3.413^{***}	-3.416^{***}	-3.406^{***}	-3.398^{***}	-3.433^{***}
	(0.032)	(0.033)	(0.032)	(0.033)	(0.032)	(0.033)
Market Concentration	0.173	0.254				0.112
	(0.157)	(0.172)				(0.158)
Buyer Concentration	-0.231^{**}	-0.260^{**}				0.115
	(0.105)	(0.108)				(0.104)
Seller Concentration	-0.051	-0.227				-0.072
	(0.098)	(0.110)				(0.100)
Number Traders	-0.620^{***}		-0.534^{**}			-0.516^{**}
	(0.207)		(0.216)			(0.217)
Number Buyers	0.175		0.356^{***}			0.365^{***}
	(0.123)		(0.131)			(0.135)
Number Sellers	0.674^{***}		0.447^{***}			0.367^{***}
	(0.133)		(0.135)			(0.137)
Client Buy Share	0.279^{***}			0.093^{***}		0.151^{***}
	(0.032)			(0.028)		(0.033)
Cum. Client Buy Share	-0.047			-0.014		-0.039
	(0.032)			(0.029)		(0.033)
Volume	0.002				-0.015	-0.016
	(0.027)				(0.039)	(0.040)
Number Traders	0.111^{***}				0.173^{***}	0.060
	(0.030)				(0.032)	(0.041)
Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	3.00	5.14	5.65	4.87	5.25	5.86
N	33624	33624	33624	33624	33624	33624
	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-3.403^{***}	-3.415^{***}	-3.418^{***}	-3.361^{***}	-3.380^{***}	-3.442^{***}
	(0.069)	(0.072)	(0.069)	(0.067)	(0.067)	(0.070)
Market Concentration	-0.045	0.125				-0.040
	(0.434)	(0.518)				(0.429)
Buyer Concentration	-0.272	-0.428				-0.229
	(0.307)	(0.339)				(0.299)
Seller Concentration	0.096	-0.121				0.060
	(0.224)	(0.291)				(0.225)
Number Traders	-0.695		-0.590			-0.619
	(0.529)		(0.543)			(0.552)
Number Buyers	0.381		0.583^{*}			0.471
	(0.322)		(0.338)			(0.350)
Number Sellers	0.548^{*}		0.389			0.379
	(0.310)		(0.314)			(0.324)
Client Buy Share	0.150^{**}			0.045		0.100
	(0.071)			(0.055)		(0.073)
Cum. Client Buy Share	-0.058			0.011		-0.030
	(0.080)			(0.064)		(0.080)
Volume	0.089^{*}				0.071	0.077
	(0.046)				(0.043)	(0.050)
Number Trades	0.081				0.199^{***}	0.055
	(0.061)				(0.048)	(0.070)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	3.67	3.77	4.70	2.77	4.00	5.08
N	7200	7200	7200	7200	7200	7200

Table 4: Predictive regression results for negative price dislocation events

This table presents results from predictive logistic regressions for negative price dislocation events on one-period lagged LOB variables, EMIR variables, and one-period lagged returns. We consider a window starting 5 minutes before the price dislocation event occurs and ending with the price dislocation event. We vary the slackness parameter by setting $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. *Controls* indicate that the specification controls for both the LOB variables as well as the one-period lagged return. Robust standard errors are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-3.297^{***}	-3.331^{***}	-3.343^{***}	-3.321^{***}	-3.330^{***}	-3.349^{***}
	(0.030)	(0.031)	(0.030)	(0.031)	(0.030)	(0.031)
Market Concentration	0.172	0.327^{**}				0.149
	(0.126)	(0.133)				(0.130)
Buyer Concentration	−0.091	-0.278^{***}				−0.127
	(0.080)	(0.087)				(0.083)
Seller Concentration	−0.112	-0.233^{***}				−0.060
	(0.089)	(0.090)				(0.091)
Number Traders	-0.559^{***}		-0.573^{**}			-0.567^{**}
	(0.214)		(0.225)			(0.221)
Number Buyers	0.472^{***}		0.336^{**}			0.256^{*}
	(0.135)		(0.134)			(0.135)
Number Sellers	0.281^{**}		0.511^{***}			0.460^{***}
	(0.126)		(0.136)			(0.139)
Client Buy Share	-0.176^{***}			−0.028		-0.080^{***}
	(0.030)			(0.026)		(0.031)
Cum. Client Buy Share	−0.011			0.008		−0.012
	(0.031)			(0.029)		(0.031)
Volume	0.006				−0.024	−0.018
	(0.013)				(0.045)	(0.041)
Number Trades	0.149^{***}				0.222^{***}	0.143^{***}
	(0.025)				(0.025)	(0.031)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	2.28	3.25	3.94	2.99	3.88	4.20
N	34436	34436	34436	34436	34436	34436
	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-3.363^{***}	-3.381^{***}	-3.410^{***}	-3.378^{***}	-3.389^{***}	-3.414^{***}
	(0.064)	(0.067)	(0.067)	(0.067)	(0.066)	(0.067)
Market Concentration	−0.115	0.030				−0.159
	(0.269)	(0.312)				(0.268)
Buyer Concentration	0.038	−0.122				0.028
	(0.170)	(0.193)				(0.171)
Seller Concentration	0.155	−0.014				0.214
	(0.186)	(0.203)				(0.182)
Number Traders	−0.722		−0.758			-0.824^{*}
	(0.499)		(0.502)			(0.499)
Number Buyers	0.529^{*}		0.307			0.297
	(0.300)		(0.272)			(0.281)
Number Sellers	0.448		0.764^{**}			0.779^{**}
	(0.285)		(0.318)			(0.323)
Client Buy Share	−0.096			0.036		−0.012
	(0.059)			(0.052)		(0.059)
Cum. Client Buy Share	−0.007			0.010		0.000
	(0.074)			(0.065)		(0.074)
Volume	0.005				−0.012	−0.060
	(0.070)				(0.070)	(0.085)
Number Trades	0.136^{**}				0.241^{***}	0.156^{**}
	(0.059)				(0.062)	(0.070)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	2.43	2.84	4.11	2.75	3.89	4.43
N	7555	7555	7555	7555	7555	7555

- Headline: "Bundesbank chief calls for German tax cuts to boost investment"
- Standfirst: "...The German economy grew 0.2 percent in the first three months of the year compared with the previous quarter. GDP fell 0.3 percent last year, making it the worst-performing major economy..."

Our sample period spans from January 2022 to July 2024, covering about two and half years. We have in total of 35,269 (12,301) news articles for the DAX (ESX) index, averaging about 267 (93) per week.

Using ChatGPT to predict crash risk

After obtaining the news articles, we ask ChatGPT to evaluate the likelihood that a particular news article is associated with an increase in the crash risk of the stock index (DAX or ESX). Specifically, we use the following prompt:

- System prompt (used to set the context): "You are a financial expert with experience in predicting stock market crashes."
- The user prompt (used to feed the specific question): "Using a scale from 1 to 10, how likely will the news increase the crash risk of the {index} stock index in the next three months? Provide your evaluation as a single number in the first line and then elaborate with one short and concise sentence on the next line. The news headline is : {headline} and the news standfirst is: {standfirst}"

Our prompt follows the general structure of Lopez-Lira and Tang (2023) but has two differences. First, instead of asking ChatGPT to provide a binary answer "Yes" or "No", we ask it to evaluate the news based on a scale from 1 to 10. Second, we not only ask ChatGPT to evaluate the headline but also the standfirst, which might contain useful information in additional to the headline. We expect these two variations help to produce more robust evaluation results.

Using the same news article above as an example, we obtain the following responses from ChatGPT when asking it to evaluate it:

- Evaluation: 4
- Explanation: "Tax cuts could stimulate economic activity and growth, potentially offsetting the recent weak performance of the German economy, and therefore decreasing the crash risk of the DAX stock index in the near term."

Table 5 reports summary statistics of the evaluation scores for news relevant for the two indices respectively.

Note that different versions of ChatGPT use different training samples. To avoid look-ahead bias, we use ChatGPT 4 to evaluate the news. The training data of ChatGPT 4 stops in September 2021 and our news start in 2022. With this we ensure that ChatGPT only uses past information to evaluate a news article, not future information.

Table 5: Summary Statistics of the news evaluations by ChatGPT.

Each news article is scored on a scale from 1 to 10 as to the likelihood that it will increase the crash risk of the corresponding stock index.The sample period cover from January 2022 to July 2024.

Index	N	Mean	SD	Min	Q25	Q50	Q75	Max
DAX	12301	5.37	1.94	1.0	4.0	6.0	7.0	9.0
ESX	35269	5.59	1.89	1.0	4.0	6.0	7.0	9.5

Constructing the fragility index

We then construct a fragility index for the DAX and ESX stock index, respectively, based on ChatGPT's evaluations of the relevant news articles. Specifically, for each week, we calculate the average evaluation across all news items and use it as the fragility measure. Recall that we ask ChatGPT to judge the likelihood that the news article is associated with an increase in the crash risk based on a scale from 1 to 10. Therefore, the larger the average evaluation or the fragility measure is, the higher crash risk is. Figure 5 plots our fragility index for the ESX and DAX stock market index over our sample period. It shows that the fragility index or the crash risk of both indices is elevated during 2022 and drops starting from 2023.

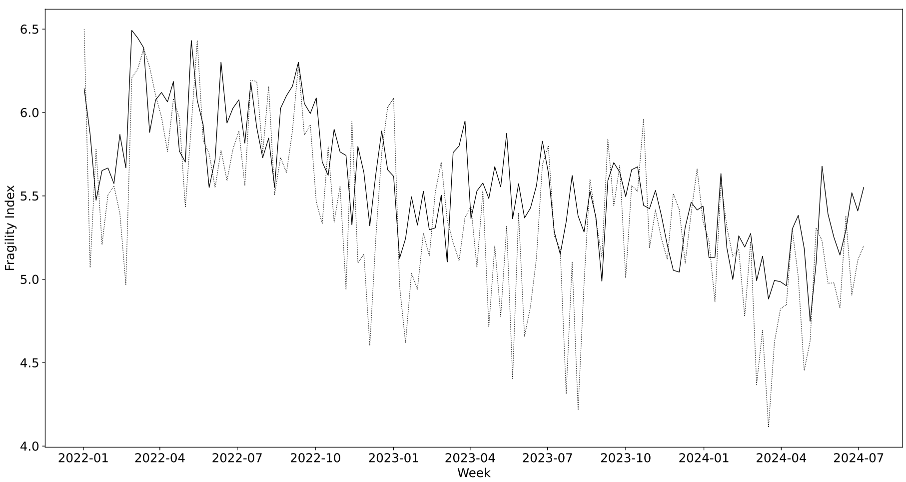
5 Conclusions

We propose a methodology to identify price dislocations in financial markets taking into account market volatility. Using trade repository data on EuroSTOXX 50 index futures, we analyze trading in the vicinity of price dislocation events.

Our results show that, on average, price dislocation events are accompanied by an increase in the number of traders on both sides of the market, as well as a decrease in market concentration. Trading volumes increase sharply and the order imbalance increases in the direction of the price dislocation. However, conditional on a price dislocation occurring, the size of the dislocation increases in the extent to which the liquidity-supplying side of the market becomes concentrated.

Figure 5: Fragility index for ESX and DAX

We construct a weekly fragility index for the ESX (solid line) and DAX (dotted line) stock market index, respectively, based on ChatGPT's evaluations of the relevant news articles from the Financial Times. Each news article is scored on a scale from 1 to 10 as to the likelihood that it is associated with an increase in the crash risk of the corresponding stock index. Then we calculate the average evaluation of all news articles within the week and use it as the fragility measure. The sample period covers January 2022 to July 2024.



References

Anand, Amber and Kumar Venkataraman (2016). “Market Conditions, Fragility, and the Economics of Market Making”. In: *Journal of Financial Economics* 121.2, pp. 327–349.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens (2001). “The Distribution of Realized Stock Return Volatility”. In: *Journal of Financial Economics* 61.1, pp. 43–76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega (2007). “Real-time Price Discovery in Global Stock, Bond and Foreign Exchange Markets”. In: *Journal of International Economics* 73.2, pp. 251–277.

Andersen, Torben G., Viktor Todorov, and Bo Zhou (2023). *Real-Time Detection of Local No-Arbitrage Violations.* Working Paper.

Bellia, Mario, Kim Christensen, Aleksey Kolokolov, Loriana Pelizzon, and Roberto Renò (2024). *Do Designated Market Makers Provide Liquidity During Extreme Price Movements?* SAFE Working Paper 270.

Bessembinder, Hendrik, Allen Carrion, Laura Tuttle, and Kumar Venkataraman (2016). “Liquidity, Resiliency and Market Quality Around Predictable Trades: Theory and Evidence”. In: *Journal of Financial Economics* 121.1, pp. 142–166.

Biais, Bruno, Fany Declerck, and Sophie Moinas (2016). *Who Supplies Liquidity, How and When?* Working Paper.

Bogousslavsky, Vincent (2021). “The Cross-section of Intraday and Overnight Returns”. In: *Journal of Financial Economics* 141.1, pp. 172–194.

Brogaard, Jonathan, Allen Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov (2018). “High Frequency Trading and Extreme Price Movements”. In: *Journal of Financial Economics* 128.2, pp. 253–265.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2019). “Price Discovery without Trading: Evidence from Limit Orders”. In: *Journal of Finance* 74.4, pp. 1621–1658.

Capponi, Agostino and Shihao Yu (2024). *Price Discovery in the Machine Learning Age.* Working Paper.

Cespa, Giovanni and Xavier Vives (2017). *High Frequency Trading and Fragility.* ECB Working Paper 2020.

Czech, Robert, Shiyang Huang, Dong Lou, and Tianyu Wang (2021). “Informed Trading in Government Bond Markets”. In: *Journal of Financial Economics* 142.3, pp. 1253–1274.

European Systemic Risk Board (June 2020). *Liquidity Risks Arising from Margin Calls.* Report.

Hendershott, Terrence and Albert J. Menkveld (2014). “Price Pressures”. In: *Journal of Financial Economics* 114.3, pp. 405–423.

Jurkatis, Simon, Andreas Schrimpf, Karamfil Todorov, and Nicholas Vause (2022). *Relationship Discounts in Corporate Bond Trading.* Working Paper.

Kaiser, Henry F (1958). “The Varimax Criterion for Analytic Rotation in Factor Analysis”. In: *Psychometrika* 23.3, pp. 187–200.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun (2017). "The Flash Crash: High-Frequency Trading in an Electronic Market". In: *Journal of Finance* 72.3, pp. 967–998.

Lopez-Lira, Alejandro and Yuehua Tang (2023). *Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models*. Working Paper.

Menkveld, Albert J. and Ion Lucas Saru (2024). *Who Knows? Information Differences Between Trader Types*. Unpublished Working Paper.

Menkveld, Albert J and Bart Zhou Yueshen (2019). "The Flash Crash: A Cautionary Tale About Highly Fragmented Markets". In: *Management Science* 65.10, pp. 4470–4488.

Montgomery, Douglas C (2019). *Introduction to Statistical Quality Control*. John Wiley & Sons.

Page, Ewan S (1954). "Continuous Inspection Schemes". In: *Biometrika* 41.1/2, pp. 100–115.

Pasquariello, Paolo (2014). "Financial Market Dislocations". In: *The Review of Financial Studies* 27.6, pp. 1868–1914.

Pinter, Gabor (2022). *An Anatomy of the 2022 Gilt Market Crisis*. Working Paper.

van Kervel, Vincent and Albert J. Menkveld (2019). "High-Frequency Trading around Large Institutional Orders". In: *Journal of Finance* 74.3, pp. 1091–1137.

A Data Cleaning

Our analysis relies on a set of variables based on the trade repositories from EMIR. In this section, we describe our data cleaning procedure. Starting from the set of all trade records, we drop any exact duplicate observations. Next, we only keep trades that correspond to a new position or a change in an existing position of the respective reporting counterparty. In addition, to make sure that the reporting information is reliable, we only keep trade records that report a positive transaction price. Table 6 summarizes the number of trade records in the raw data as well as after we implement our data cleaning steps. Moreover, comparing dislocation days with non-dislocation days shows that our data cleaning procedure does not remove more observations on days with price dislocation events than on days with no recorded price dislocations.

We merge the order book data from BMLL with the trade repository data from EMIR based on the International Securities Identification Number (ISIN).

Table 6: Trade records after data cleaning steps

This table summarizes the total number of trade records in the raw data as well as after we implement our data cleaning steps. In addition, we report the percentage of trade records that our data cleaning procedure drops.

Instrument	Raw Data	After Cleaning	Percentage Dropped
Panel A: Price Dislocation Days			
FESX	430,157,907	429,590,471	0.132%
Panel B: Non-Dislocation Days			
FESX	361,898,194	361,088,859	0.224%

B Additional Results

In this section, we report additional results that are omitted in the main section for brevity. Table 7 presents summary statistics on the coefficients from estimating a GARCH model for our SPC measure (see Section 2 for details).

Tables 8 and 9 present results for contemporaneous logistic regressions in a 10-minute interval around positive and negative price dislocations, respectively.

Figure 6 exhibits the loadings of the EMIR and LOB variables on the first three factors extracted using an exploratory factor analysis with varimax rotation (Kaiser, 1958).

Figure 7 compares the distribution of the price dislocation events identified using our SPC measure with the distribution of extreme price movements following Brogaard et al. (2018), and conditional volatility forecasts over the sample period.

Figure 8 exhibits the number of selling counterparties around transitory and permanent positive price dislocation events and Figure 9 shows number of buying counterparties around transitory and permanent negative price dislocation events.

Figure 10 plots the share of client buy orders in all client orders for 10-minute intervals around price dislocation events.

Table 7: GARCH estimation results

This table presents summary statistics on the coefficient from estimating a GARCH model for daily volatilities using an expanding window.

	Constant	ARCH	GARCH	Offset	Degrees of Freedom
Mean	0.165	0.113	0.848	0.058	5.711
Median	0.142	0.116	0.853	0.064	5.618
Standard Deviation	0.156	0.021	0.065	0.044	1.489
5% Percentile	0.109	0.095	0.829	−0.010	4.664
95% Percentile	0.194	0.131	0.881	0.119	6.082
N	999	999	999	999	999

Table 8: Logistic regression results for positive price dislocation events

This table presents logistic regression results of an indicator for positive price dislocation events on LOB variables and EMIR variables for a window ranging from 5 minutes before to 5 minutes after a price dislocation takes place. We vary the slackness parameter by setting $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. Robust standard errors are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-4.762^{***}	-5.126^{***}	-5.134^{***}	-5.185^{***}	-5.255^{***}	-5.460^{***}
	(0.046)	(0.050)	(0.050)	(0.050)	(0.061)	(0.063)
Market Concentration	1.013^{***}	0.964^{***}				0.715^{***}
	(0.086)	(0.111)				(0.124)
Buyer Concentration	-0.833^{***}	-0.645^{***}				-0.414^{***}
	(0.067)	(0.087)				(0.093)
Seller Concentration	-0.456^{***}	-0.643^{***}				-0.270^{***}
	(0.067)	(0.095)				(0.099)
Number Traders	-1.349^{***}		-1.394^{***}			-1.175^{***}
	(0.194)		(0.261)			(0.267)
Number Buyers	0.283^{***}		0.502^{***}			0.623^{***}
	(0.104)		(0.145)			(0.151)
Number Sellers	1.627^{***}		1.195^{***}			0.975^{***}
	(0.131)		(0.171)			(0.176)
Client Buy Share	0.993^{***}			0.306^{***}		0.383^{***}
	(0.031)			(0.033)		(0.045)
Cum. Client Buy Share	-0.136^{***}			-0.049		-0.086^{*}
	(0.040)			(0.040)		(0.046)
Volume	0.007				-0.884^{***}	-0.827^{***}
	(0.027)				(0.136)	(0.134)
Number Trades	0.234^{***}				0.569^{***}	0.428^{***}
	(0.027)				(0.069)	(0.063)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2(%)	19.73	39.07	39.75	39.13	40.79	42.12
N	61670	61670	61670	61670	61670	61670

	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-4.989^{***}	-5.377^{***}	-5.337^{***}	-5.327^{***}	-5.405^{***}	-5.666^{***}
	(0.108)	(0.130)	(0.126)	(0.118)	(0.132)	(0.150)
Market Concentration	1.470^{***}	1.370^{***}				0.907^{***}
	(0.316)	(0.280)				(0.310)
Buyer Concentration	-1.209^{***}	-0.911^{***}				-0.475
	(0.185)	(0.261)				(0.297)
Seller Concentration	-0.745^{***}	-1.008^{***}				-0.538^{**}
	(0.249)	(0.232)				(0.220)
Number Traders	-1.935^{***}		-2.417^{***}			-2.271^{***}
	(0.491)		(0.673)			(0.633)
Number Buyers	0.633^{**}		1.429^{***}			1.554^{***}
	(0.272)		(0.373)			(0.357)
Number Sellers	1.810^{***}		1.338^{***}			1.130^{***}
	(0.318)		(0.420)			(0.412)
Client Buy Share	1.016^{***}			0.180^{**}		0.377^{***}
	(0.070)			(0.075)		(0.104)
Cum. Client Buy Share	-0.134			-0.036		-0.120
	(0.110)			(0.090)		(0.120)
Volume	0.171^{***}				-0.578^{***}	-0.662^{***}
	(0.059)				(0.210)	(0.210)
Number Trades	0.230^{***}				0.526^{***}	0.471^{***}
	(0.058)				(0.121)	(0.118)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	24.51	45.51	46.27	45.08	46.61	48.18
N	13609	13609	13609	13609	13609	13609

Table 9: Logistic regression results for negative price dislocation events

This table presents logistic regression results of an indicator for negative price dislocation events on LOB variables and EMIR variables for a window ranging from 5 minutes before to 5 minutes after a price dislocation takes place. We vary the slackness parameter by setting $k = 3\sigma$ and $k = 6\sigma$. All explanatory variables are standardized to have a mean of zero and a standard deviation of one. Robust standard errors are reported in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, *** denotes significance at the 1% level.

	Panel A: $k = 3\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-4.874^{***}	-4.970^{***}	-4.971^{***}	-4.978^{***}	-5.071^{***}	-5.365^{***}
	(0.048)	(0.048)	(0.047)	(0.047)	(0.055)	(0.060)
Market Concentration	1.100^{***}	1.117^{***}				0.804^{***}
	(0.085)	(0.104)				(0.112)
Buyer Concentration	-0.485^{***}	-0.777^{***}				-0.278^{***}
	(0.065)	(0.099)				(0.102)
Seller Concentration	-1.024^{***}	-0.924^{***}				-0.685^{***}
	(0.066)	(0.081)				(0.080)
Number Traders	-1.279^{***}		-1.328^{***}			-0.727^{***}
	(0.201)		(0.261)			(0.257)
Number Buyers	1.668^{***}		1.274^{***}			0.801^{***}
	(0.131)		(0.171)			(0.176)
Number Sellers	0.130		0.494^{***}			0.355^{**}
	(0.109)		(0.136)			(0.144)
Client Buy Share	-1.052^{***}			-0.348^{***}		-0.497^{***}
	(0.032)			(0.028)		(0.045)
Cum. Client Buy Share	-0.053			0.022		-0.059
	(0.039)			(0.039)		(0.045)
Volume	-0.141^{***}				-1.539^{***}	-1.319^{***}
	(0.054)				(0.181)	(0.216)
Number Traders	0.371^{***}				0.711^{***}	0.531^{***}
	(0.035)				(0.051)	(0.057)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	21.94	35.39	36.60	35.18	38.21	40.15
N	63455	63455	63455	63455	63455	63455
	Panel B: $k = 6\sigma$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-5.177^{***}	-5.287^{***}	-5.300^{***}	-5.268^{***}	-5.392^{***}	-5.626^{***}
	(0.118)	(0.112)	(0.105)	(0.107)	(0.123)	(0.136)
Market Concentration	1.095^{***}	0.993^{***}				0.635^{***}
	(0.171)	(0.218)				(0.222)
Buyer Concentration	-0.445^{***}	-0.696^{***}				-0.169
	(0.137)	(0.213)				(0.198)
Seller Concentration	-0.960^{***}	-0.755^{***}				-0.392
	(0.175)	(0.220)				(0.246)
Number Traders	-1.383^{***}		-0.388			-0.033
	(0.467)		(0.682)			(0.502)
Number Buyers	1.953^{***}		0.328			0.271
	(0.333)		(0.430)			(0.307)
Number Sellers	0.020		0.575			0.338
	(0.233)		(0.365)			(0.327)
Client Buy Share	-1.011^{***}			-0.172^{***}		-0.347^{***}
	(0.074)			(0.066)		(0.099)
Cum. Client Buy Share	-0.063			0.096		-0.002
	(0.107)			(0.097)		(0.134)
Volume	-0.028				-0.722^{***}	-0.799^{***}
	(0.069)				(0.179)	(0.185)
Number Trades	0.432^{***}				0.766^{***}	0.608^{***}
	(0.065)				(0.099)	(0.098)
LOB Controls	No	Yes	Yes	Yes	Yes	Yes
Pseudo R^2 (%)	27.60	42.87	44.20	42.55	46.26	47.60
N	14675	14675	14675	14675	14675	14675

Figure 6: Exploratory factor analysis of variables

This figure plots the factor loadings of the variables derived from the order book data and the transaction records from EMIR on their first three factors from an exploratory factor analysis. The factors are rotated using the varimax rotation (Kaiser, 1958) and computed based on a window around price dislocation event, starting 5 minutes before and ending 5 minutes after price dislocation events occur. We use a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. Order book state variables are denoted by circles, other order book variables are denoted by downward pointing triangles, concentration measures from EMIR are denoted by upward pointing triangles, and other variables derived from EMIR are denoted by squares.

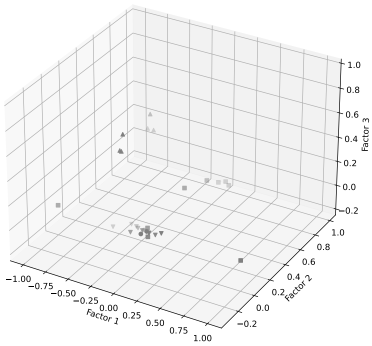
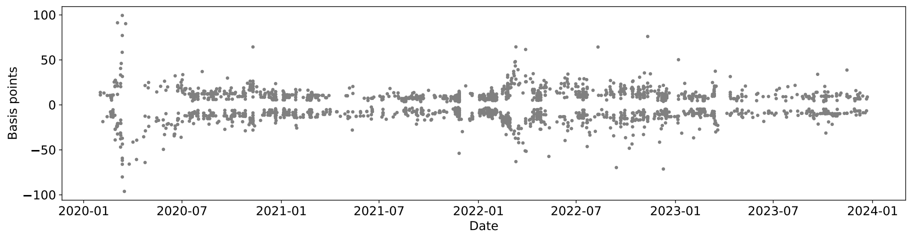
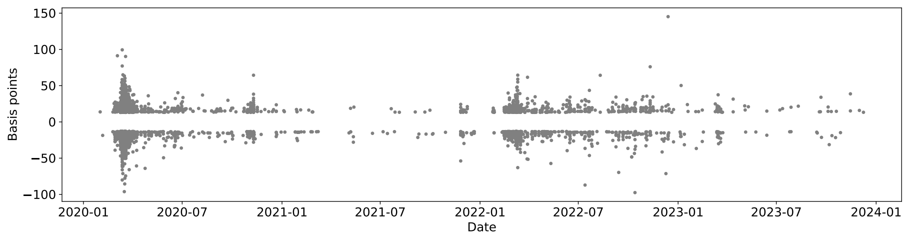


Figure 7: Price dislocation events, extreme price movements, and conditional volatility forecasts

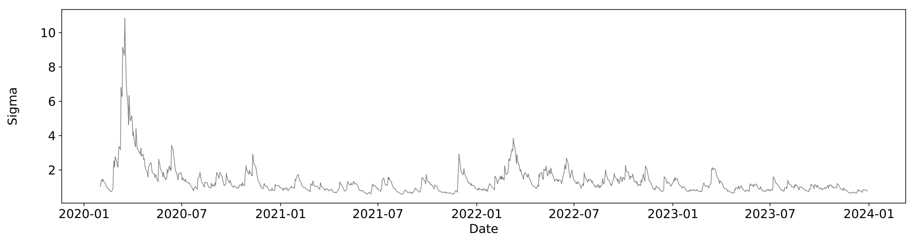
This figure plots price dislocation events for a slackness of $k = 3\sigma$ and a control threshold of $h = 5\sigma$ in Panel (a). For each dislocation event, we plot the corresponding 10-second return in basis points. Panel (b) shows extreme price movements according to the definition of Brogaard et al. (2018): Absolute returns that are above 99.9th percentile of the absolute return distribution. Again, we plot the corresponding 10-second return in basis points. Panel (c) shows one-step ahead conditional volatility forecasts from a GARCH model.



(a) Price dislocations over time



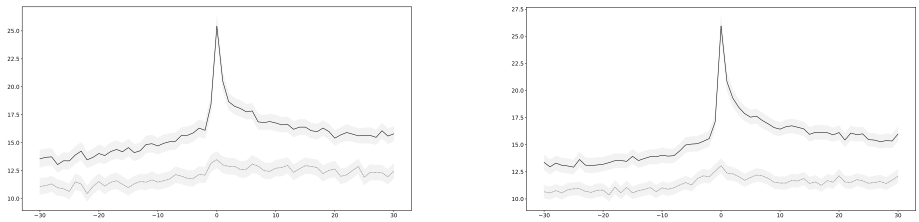
(b) Extreme price movements over time



(c) Conditional volatility forecasts over time

Figure 8: Selling counterparties in the vicinity of positive price dislocation events

The figure plots the number of selling counterparties for a 10-minute interval around positive price dislocation events with a slackness parameter of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. We distinguish between transitory and permanent price dislocation events. Transitory price dislocation events revert by more than 2/3 at the end of a 30-minute period and permanent price dislocation events revert by less than 1/3 after 30 minutes. Solid lines refer to the treatment sample and dotted lines to the placebo sample, with shaded areas being the corresponding 95% confidence intervals. We plot the variables for each 10-second interval, starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.

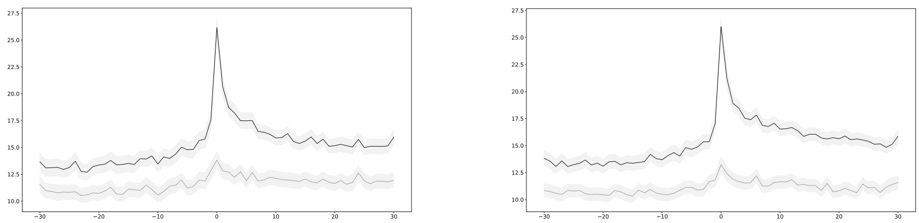


(a) Transitory price dislocations

(b) Permanent price dislocations

Figure 9: Buying counterparties in the vicinity of negative price dislocation events

The figure plots the number of buying counterparties for a 10-minute interval around negative price dislocation events with a slackness parameter of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. We distinguish between transitory and permanent price dislocation events. Transitory price dislocation events revert by more than 2/3 at the end of a 30-minute period and permanent price dislocation events revert by less than 1/3 after 30 minutes. Solid lines refer to the treatment sample and dotted lines to the placebo sample, with shaded areas being the corresponding 95% confidence intervals. We plot the variables for each 10-second interval, starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.

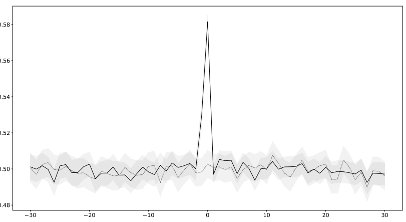


(a) Transitory price dislocations

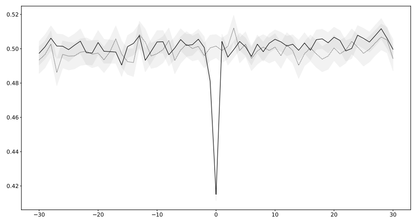
(b) Permanent price dislocations

Figure 10: Client buy share in the vicinity of price dislocation events

This figure plots the share of client buy orders in all client orders for a 10-minute interval around price dislocation events with a slackness parameter of $k = 3\sigma$ and a control threshold of $h = 5\sigma$. Solid lines refer to the treatment sample and dotted lines to the placebo sample, with shaded areas being the corresponding 95% confidence intervals. We plot the client buy share for each 10-second interval, starting 5 minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.



(a) Positive dislocation events



(b) Negative dislocation events

C March 2020

European equity derivatives markets were subject to margin call events in March 2020 (ESRB, 2020). In this Section, we zoom-in on price dislocations occurring in March 2020 and investigate whether the average patterns that we identify in the overall sample also hold in March 2020.

We consider only the largest price dislocations with a slackness of $k = 6\sigma$ and a control threshold of $h = 5\sigma$. In total, we identify 48 positive price dislocations events and 36 negative price dislocation events in March 2020. The average return of positive price dislocation events is 129.33bps and -123.46bps for negative price dislocation events.

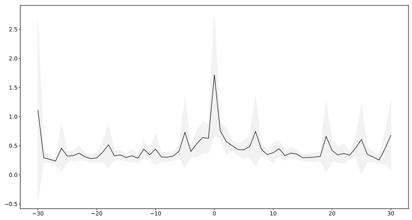
For positive price dislocations we find a decrease in buyer and seller concentration in the overall sample. In contrast, we do not find evidence for a change in either buyer or seller concentration for positive price dislocation events in March 2020. In addition, we do not find evidence for a change in the number of buying or selling counterparties in the vicinity of price dislocations. The average number of buying and selling counterparties is low, with on average 9.01 buyers and 8.95 sellers being active in a 10-minute window around positive price dislocation events.

The results for negative price dislocation events are comparable. Again, we do not find evidence for a decrease in buyer or seller concentration around negative price dislocation events in March 2020. Similarly, the number of buying and selling counterparties does not change significantly in the vicinity of price dislocation events, averaging 10.42 buying counterparties and 10.26 selling counterparties in a 10-minute window around the price dislocation events. These results suggest that dislocations occurring in March 2020 differ from the average price dislocation in our sample. At the same time, these findings alleviate the concern that our results are driven by market movements in March 2020.

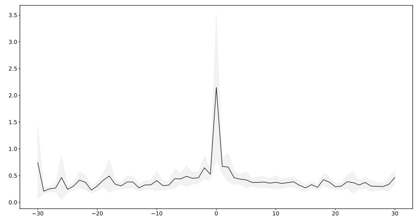
D First-time Buyers and First-time Sellers

Figure 11: First-time buyers and first-time sellers in the vicinity of positive price dislocation events

This figure plots the number of first-time buyers (counterparties that act as buyers for the first time during the trading day) and first-time sellers (counterparties that act as sellers for the first time during the trading day) around positive price dislocation events, together with 95% confidence intervals. We set the slackness $k = 3\sigma$ and use a control threshold of $h = 5\sigma$. We plot the number of first-time buyers and first-time sellers for each 10-second interval, starting 5-minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.



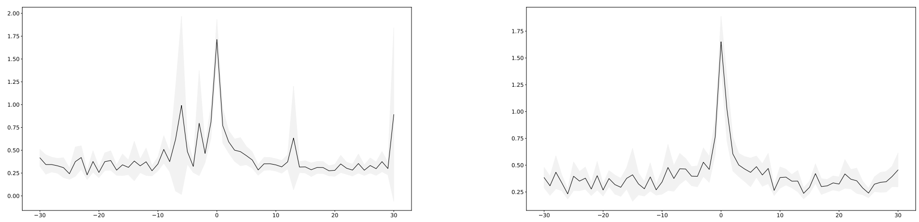
(a) First-time buyers



(b) First-time sellers

Figure 12: First-time buyers and first-time sellers in the vicinity of negative price dislocation events

This figure plots the number of first-time buyers (counterparties that act as buyers for the first time during the trading day) and first-time sellers (counterparties that act as sellers for the first time during the trading day) around negative price dislocation events, together with 95% confidence intervals. We set the slackness $k = 3\sigma$ and use a control threshold of $h = 5\sigma$. We plot the number of first-time buyers and first-time sellers for each 10-second interval, starting 5-minutes before the price dislocation takes place and ending 5 minutes after the price dislocation takes place.



(a) First-time buyers (b) First-time sellers

Imprint and acknowledgements

This project was carried out within the Alberto Giovannini Programme for Data Science, generously provided by the European Systemic Risk Board (ESRB). We benefited greatly from insightful feedback and extensive data support from various members of the ESRB Secretariat, including Marco D'Errico, Francesco Mazzaferro, and Tuomas Peltonen.
The views expressed in this publication are those of the authors and not necessarily reflect the official stance of the ESRB/ECB, its member institutions or the institutions to which the authors are affiliated.

Albert J. Menkveld
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; email: **albertjmenkveld@gmail.com**

Ion Lucas Saru (corresponding author)
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; Tinbergen Institute, Amsterdam, The Netherlands; email: **saru.ionlucas@gmail.com**

Shihao Yu
Singapore Management University, Singapore, Singapore; email: **shihaoyu@smu.edu.sg**

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website **www.esrb.europa.eu**

For specific terminology please refer to the **ESRB glossary** (available in English only).

PDF ISBN 978-92-9472-431-1, ISSN 2467-0677, doi:10.2849/0498883, DT-01-25-015-EN-N