FLASH CRASH GENERATION AND PREVENTION

Report prepared as part of

MASTERS OF FINANCIAL ENGINEERING PROGRAM

UNIVERSITY OF CALIFORNIA LOS ANGELES

by Edgar Adel

under supervision of

Jack Sarkissian
Managing Partner
Algostox Trading

(Fall 2023)

PROJECT OBJECTIVE

The objective of this project is to investigate the conditions necessary to generate a flash crash in the equities market, similar to the one that occurred on May 6, 2010, and other similar events. The project aims to determine various ways and strategies with which such events can be artificially generated and derive recommendations for flash crash prevention mechanisms based on the findings. To achieve the goal, we built a microstructural model that reflects the observed behaviours and studied its numerical solutions using Python. Based on the findings, we derive recommendations for flash crash prevention mechanisms.

APPROACH

We look at market crashes as events in which the market loses the balance between supply and demand. For this we used the statistical physics model that describes the behaviour of the market through the imbalance field - the difference between probabilities of trade executing at the bid or ask levels at different price levels [1]. When a security trades at its fair value, both sides are equivalent and the imbalance is zero. When a sudden repricing occurs, the imbalance profile undergoes changes to conform to the new fair value. This transition can occur smoothly, or turbulently, depending on the size and speed of the price shift. Crash generation can be achieved by managing the repricing process to drive the market into the turbulent self-sustainable transition.

INTRODUCTION AND BACKGROUND

Flash crashes are rapid and severe market downturns within a concise time frame, often just a few minutes. These events can lead to significant losses for investors and can have a destabilizing effect on financial markets. Understanding the factors contributing to flash crashes and developing effective prevention strategies is crucial for maintaining the stability and integrity of financial markets. This report explores the generation and prevention of flash crashes, focusing on three hypotheses that shed light on the underlying causes.

Flash crash examples

- May 6, 2010, flash crash
- April 23, 2013, flash crash
- Frankenshock,or Flash Crash Swiss Franc on January 15, 2015

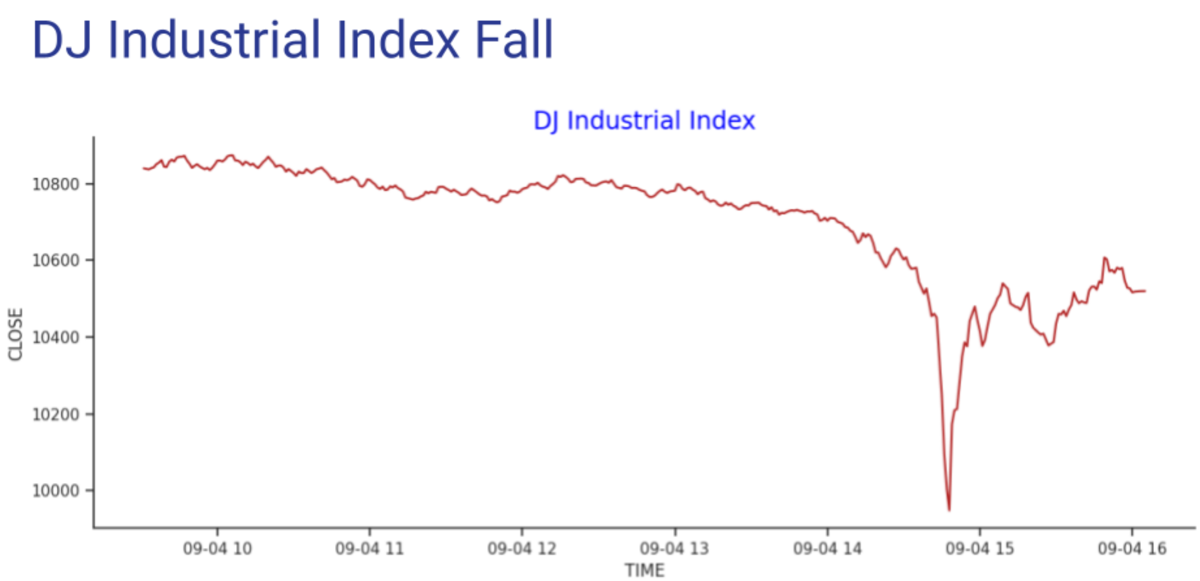
- Flash Crash of the British Pound on October 6, 2016
- Flash Crash of Japanese Yen on January 2, 2019
- Flash Crash of European Stock Markets on May 2, 2022.

The 2010 flash crash history

On May 6, 2010, a $4.1 billion trade on the New York Stock Exchange resulted in a sudden and significant loss to the Dow Jones Industrial Average, followed by a quick recovery, all within about fifteen minutes. This event, known as the 2010 Flash Crash, has been the subject of heavy research and remains a topic of dispute regarding its underlying causes.

In a few minutes Dow Jones Industrial Average lost about 8% going from 10,650 points to below 10,000 points, recovering in less than 10 minutes up 10,520 points.

To better understand the dynamics of flash crashes, we can examine the historical example of the 2010 Flash Crash that occurred on May 6th, 2010. This event can be divided into several phases:



Phase I: From the market open through about 2:32 p.m., prices were broadly declining across markets, with stock market index products sustaining losses of about 3%.
Phase II: From about 2:32 p.m. through about 2:41 p.m., the broad markets began to lose more ground, declining another 1-2%.
Phase III: In the third phase lasting only about four minutes or so, volume spiked upwards, and the broad markets plummeted a further 5-6% to reach intraday lows of 9-10%.

Phase IV: From 2:45 p.m. to about 3:00 p.m. Broad market indices recovered, while at the same time many individual securities and ETFs experienced extreme price fluctuations and traded in a disorderly fashion at prices as low as one penny or as high as $100,000.
Phase V: At about 3:00 p.m., prices of most individual securities significantly recovered, and trading resumed in a more orderly fashion.

Hypothesis A: Index Up and Hedge with Short E-mini Futures

When the index, such as the S&P 500 (SPY), experiences a sudden upward movement, traders and investors may hedge their positions by shorting E-mini futures contracts. If the market subsequently reverses, they are forced to sell their futures contracts to cover their positions. This selling pressure and margin calls can lead to a cascade of liquidations, exacerbating the flash crash.
This hypothesis suggests that the interconnectedness of index-linked instruments and derivative products can trigger flash crashes. When traders anticipate a market correction after a sharp upward movement, they may hedge their long positions with short E-mini futures. If the market reverses, a wave of sell orders can be triggered, as traders may scramble to cover the different positions, leading to a sudden and steep market decline. Margin calls may add further pressure as traders are forced to sell assets to meet their margin requirements, causing a vicious selling cycle.
During the 2010 Flash Crash, the rapid decline in index-linked products and the subsequent selling pressure triggered by margin calls likely contributed to the severity of the crash. The initial market decline in Phase I and Phase II may have prompted traders to short E-mini futures to hedge against further losses.

Hypothesis B: Imbalance in Index Components

Once Flash crashes occur, its cycle may exacerbate when a significant number of shares in a particular stock, such as Procter & Gamble (P&G), are sold rapidly. For example, the Dow Jones Industrial Average is not capitalization-balanced, meaning that when a single component like P&G experiences a sharp decline, it can disproportionately impact the entire index; this can trigger a repricing effect, leading to further selling and potentially causing a flash crash.
In this scenario, a large sell order in a single stock can cause a sudden drop in its price. Since the Dow Jones is price-weighted, the decline in this one component can lead to a significant downward adjustment in the entire index; this, in turn, prompts algorithmic and high-frequency traders to sell other assets linked to the index, creating a cascading effect that can result in a flash crash.
The 2010 Flash Crash was partially driven by the massive sell-off in individual securities, including Procter & Gamble, which experienced extreme price fluctuations. As these stocks

tumbled, the Dow Jones, being price-weighted, saw a significant decline, further fueling panic selling in other index-linked instruments.

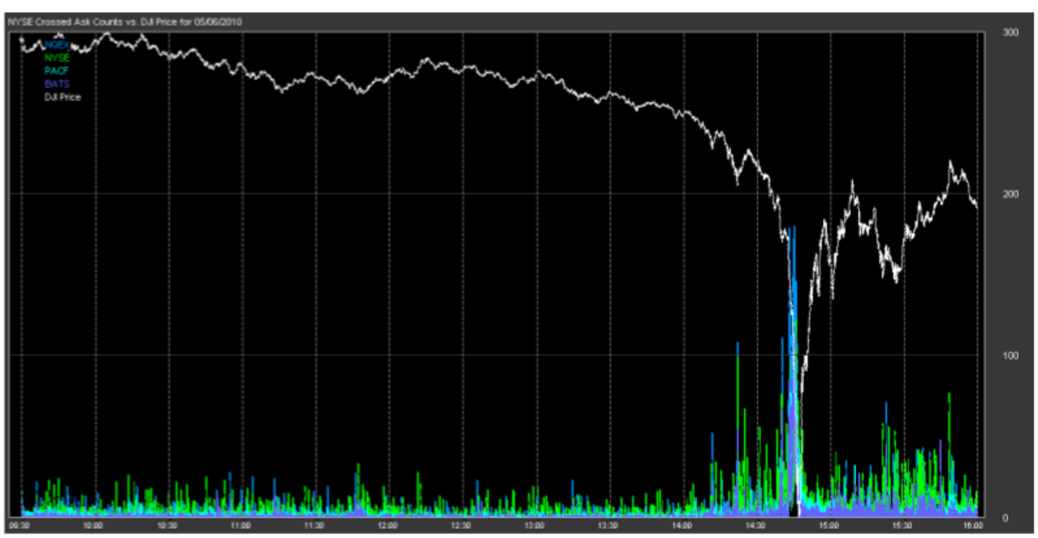
Hypothesis C: Index Arbitrage with Smaller Subsets

Also, traders can initiate flash crashes by engaging in index arbitrage, working with smaller subsets instead of the entire index. By creating an artificial index with fewer stocks, traders can exploit price discrepancies between the synthetic and traditional indexes until they converge, potentially causing dislocations in the broader market. Traditional market indices are based on the market capitalization of individual stocks or bonds. On the other hand, synthetic indices are created using derivatives and do not necessarily reflect the actual market capitalization of the underlying assets.

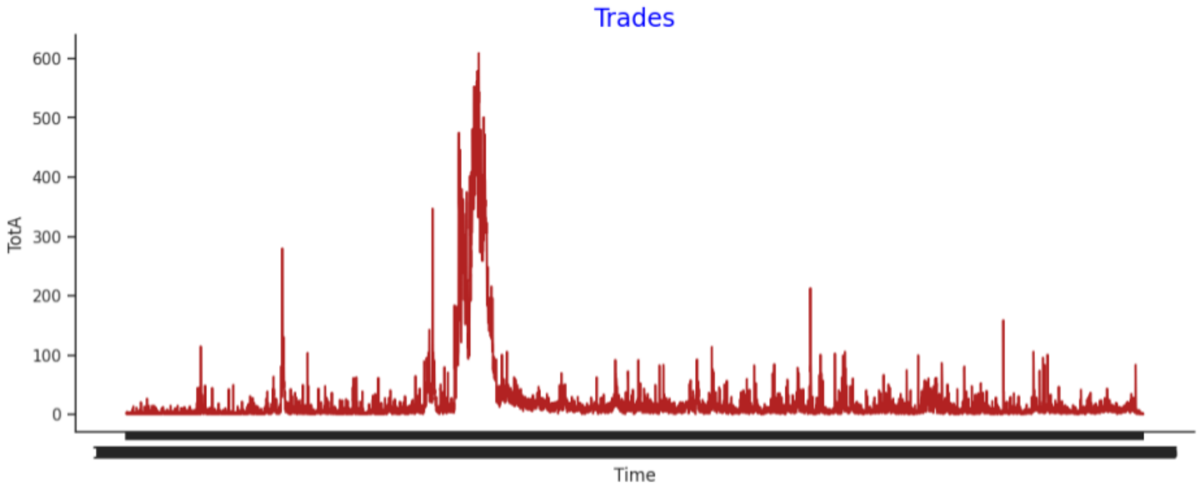
This hypothesis posits that traders may attempt to take advantage of price differences between a subset of index constituents and the broader index. This strategy involves buying and selling stocks within the subset to profit from price convergence. However, when executed on a large scale, it can create volatility and disruptions in the broader market as other traders and algorithms react to these price movements.

While the 2010 Flash Crash had elements of index arbitrage, it was mainly driven by the first two hypotheses. However, the use of smaller subsets for arbitrage purposes can introduce additional complexity and volatility into the market, especially when executed on a large scale.

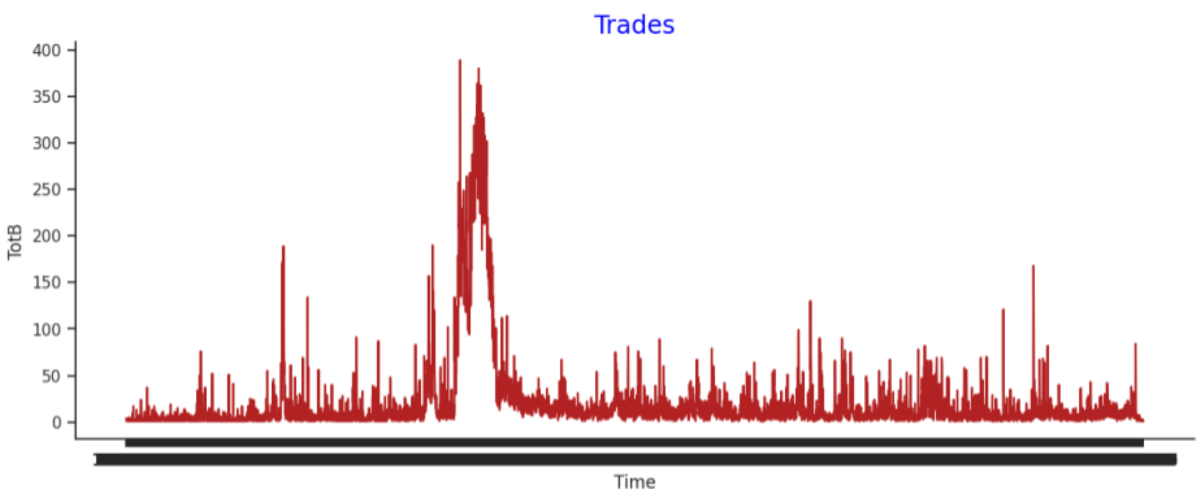
Orders vs Index



Trades per second TotA



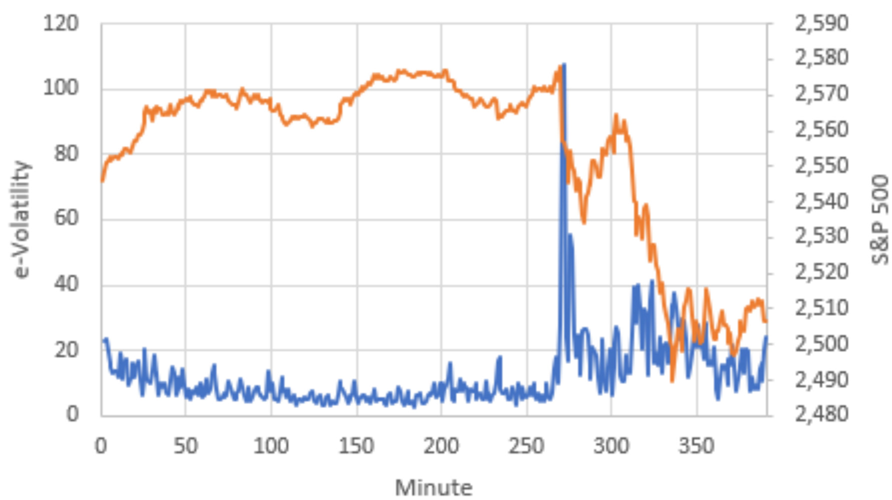
Trades per second TotB



http://www.nxcoreapi.com/20100506/FlashCrashAnalysis_Part1-3.html

OBSERVATIONS

On microstructural level market crashes, and flash crashes in particular, are characterized by a typical volatility spike followed by its gradual decrease, best measured by the ensemble volatility. Ensemble volatility or e-Volatility shows the ensemble average of returns averaged over the ensemble of index constituents. Ensemble volatility is a fast-reacting and objective measure of volatility, free of past data and future speculations developed by Sarkissian and Sebold in [2].
As the crash develops, the bid-ask spread widens and order book becomes sparse with substantially decreased order density. This is reflected in the ensemble volatility with short time-horizon, which shows the overall cost of liquidity in the market.
Picture below shows a 1-minute e-Volatility spike in a typical market crash. One can see that the e-Volatility, as well as the liquidity costs jumped from 7 bp to 100 bp (1%) within minutes, as the market went into the drawdown. This means that if someone wished to liquidate their position by taking liquidity, they would have to pay 1% no matter how small is their order. At the peak of the flash crash of 2010 the liquidity cost reached 3% even for the blue chip stocks.



MICROSTRUCTURAL MODEL

Imbalance and Fair Value

It is customary to model market crashes using the time series of security prices. However, security prices are only derivative elements to the underlying microstructural processes that are completely missed in the modeling of price time series. A more basic element of microstructural dynamics of the execution imbalance, which shows how much more the ask-price is more probable than the bid-price in trading [1, 3]:

$$I = p_{ask} - p_{bid}$$

As a function of price, the imbalance field $I(s)$ shows what imbalance would a security experience, if it were traded as price s. Following its definition, the execution imbalance can vary from -1 to 1: $-1 \leq I \leq 1$. A security that is traded at its **fair value** would experience zero imbalance since the bid and ask prices are equiprobable: $I = 0$. An overpriced security would experience negative imbalance, due to the fact that the bid-price is more probable than the ask: $I < 0$. An underpriced security would experience a positive imbalance: $I > 0$.

Unlike the balanced markets in which the fair value is well defined and quite known to market participants, during market crashes the fair value becomes highly uncertain. Technically, this means that there can be multiple points of zero imbalance, each representing a possible fair value. Imbalance dynamics with multiple zero-imbalance zones is an indication of an unstable market with a developing crash. With this in mind, let us proceed laying out the theory of balanced and further, imbalanced markets.

Balanced Market

Since the state of security is defined solely by the imbalance, so is its free energy. Taking into account that $|I| \leq 1$, :it is feasible to write the expansion of free energy as

$$F = \frac{a_2}{2} I^2 + \frac{a_4}{4} I^4 + \frac{b_2}{2}\left(\frac{\partial I}{\partial s}\right)^2$$

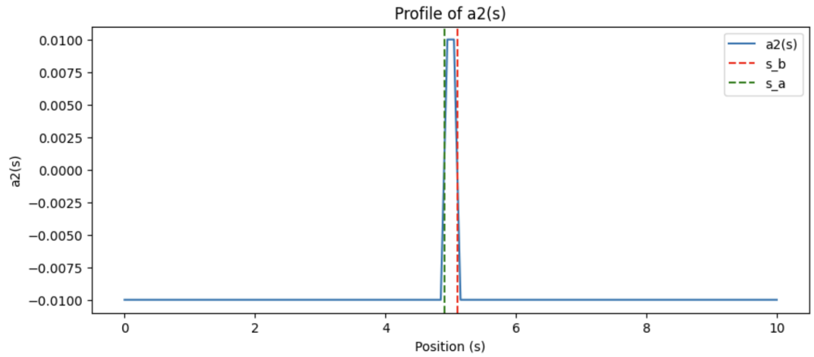
Here the odd-power terms have to be missing to observe the $s \rightarrow -s$ symmetry, since the supply and demand sides of the market are equivalent. From here, the equilibrium equation is:

$$a_2 I + a_4 I^3 - b_2 \frac{\partial^2 I}{\partial s^2} = 0$$

This equation has a physically viable analytic solution:

$$I(s) = -tanh\left(\frac{s - s_0}{\Delta s}\right)$$

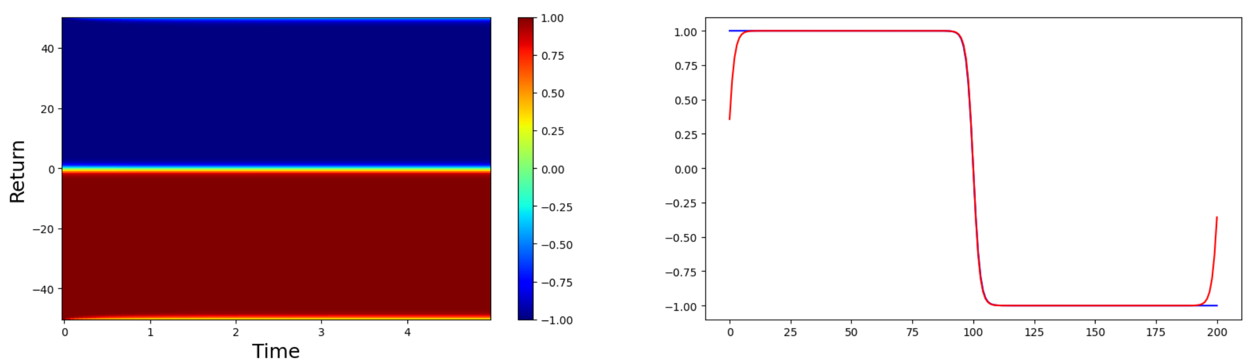
which is possible only if $a_2 = -a_4$ and $a_2 < 0$. The s_0 is the fair value and its location is determined by the structure of the a and b coefficients, in this case, a_2, as shown in the picture below.



The analytic solution can be replicated numerically by using the dynamic equation:

$$a_2 I + \ a_4 I^3 - b_2 \frac{\partial^2 I}{\partial s^2} = \ \lambda \frac{\partial I}{\partial t}$$

Under this equation any arbitrary (but reasonable) initial condition eventually evolves into the stable numeric solution which is identical to the analytic. This is a checkpoint, confirming the validity of the code.



Imbalanced Market

The free energy of the previous section corresponds to the balanced markets. In imbalanced markets free energy takes a more general form: for example, the up and down directions are no longer equivalent, and therefore the reflection symmetry is lost, which allows for the odd powers of I to participate:

$$F \ = a_1 I + \frac{a_2}{2} I^2 + \frac{a_3}{3} I^3 + \frac{a_4}{4} I^4 + \frac{b_2}{2}\left(\frac{\partial I}{\partial s}\right)^2 + \frac{\mu_2}{2}\left(\frac{\partial I}{\partial t}\right)^2$$

Here we also added the kinetic term $\frac{\mu_2}{2}\left(\frac{\partial I}{\partial t}\right)^2$, which becomes important in a dynamic process of repricing. Naturally, so satisfy the limiting case of the balanced market, we must have coefficients $a_1(s_0) = 0$ and $a_3(s_0) = 0$.

The dynamic equation corresponding to the free energy of imbalanced market is:

$$a_1 + a_2 I + \; a_4 I^3 - b_2 \frac{\partial^2 I}{\partial s^2} + \; {\tau_1}^2 \frac{\partial^2 I}{\partial t^2} + \tau_2 \frac{\partial I}{\partial t} = 0$$

where for small deviations from the fair value s_0, coefficient $a_1(s)$ is:

$$a_1(s) \sim s - s_0$$

We take it to have the following form:

$$a_1(s) = tanh\left(\frac{s - s_0}{\Delta s}\right)$$

Numerical Solution

We solve the dynamics of imbalance numerically, by split-step method after presenting the equation as a system of equations:

$$a_1 + a_2 I + \; a_4 I^3 - b_2 \frac{\partial^2 I}{\partial s^2} + \; \tau_1 \frac{\partial \Phi}{\partial t} + \frac{\tau_2}{\tau_1} \Phi = 0$$

$$\Phi = \tau_1 \frac{\partial I}{\partial t}$$

$$\Theta(t + \Delta t) = \Phi(t) - \left(\; a_1 + a_2 I + \; a_4 I^3 - b_2 \frac{\partial^2 I}{\partial s^2} \right) \frac{\Delta t}{\tau_1}$$

$$\Phi(t + \Delta t) = \Theta(t + \Delta t) \; e^{-\frac{\Delta t}{\tau_2}}$$

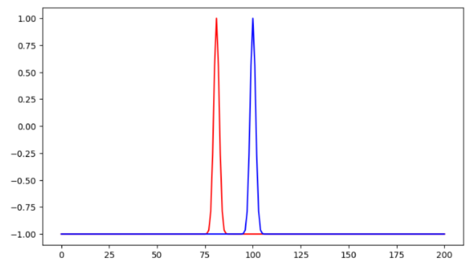
$$I(t + \Delta t) = I(t) \; + \Phi \frac{\Delta t}{\tau_1}$$

MARKET DESTABILIZATION

Seemingly, the main characteristic of a flash crash is a quick and large drawdown. This is how it appears to most people. However, this is only the superficial, external characteristic, much like a symptom in medicine. The underlying root of flash crashes is in market destabilization, in which the perception of fair value is lost or largely blurred. This characteristic is represented by the bid-ask spread of securities, which spikes during a crash. For example, in the flash crash of 2010 the spreads in securities, in which normal spreads were only 3-5 bp, would reach 3%!

To achieve such a drastic change, spread has to be destabilized and taken to an exponentially growing regime. In order to achieve it the market has to be bombarded with unidirectional sell orders taking liquidity from the buying side and adding liquidity to the selling side for a sufficiently long time. Such activity tends to increase the spread with each step and drive the price down. However, not every unidirectional flow can cause a crash. The question is: what kind of order flow will cause a flash crash and how to find its optimal conditions?

Let us see how security repricing occurs in our model. Location of the fair value is not explicitly defined in the dynamic equation. It is, however, determined by the coefficients $a_1(s)$ and $a_2(s)$, which are centered on the fair value. When security repricing occurs it is reflected in these coefficients, which simply shift to the new fair value, as shown in the picture below.

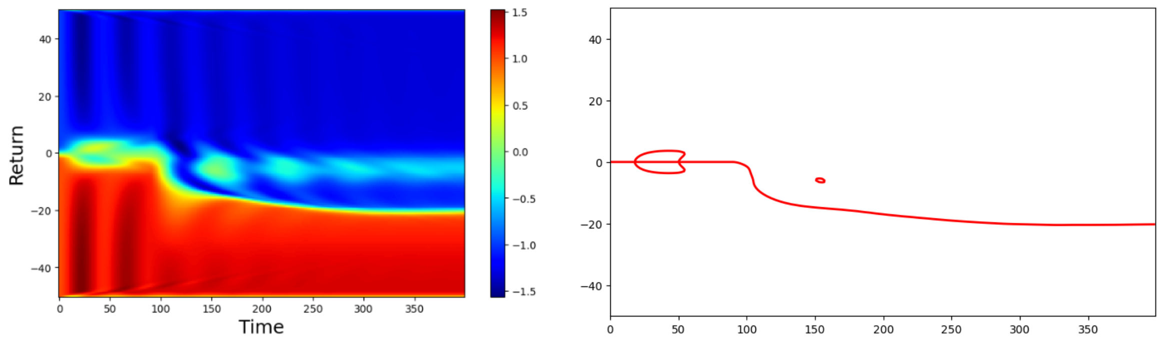


When this occurs, the profile of $I(s)$ will adjust to conform to the new settings. How this happens depends on the size and speed of the price shift.

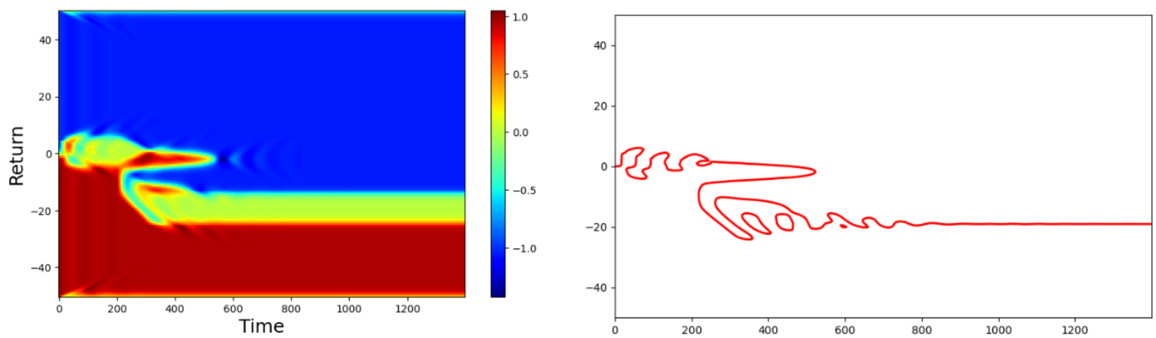
In normal, balanced markets, the $I(s)$ function is a monotonically decreasing function with only one zero. It is natural, since the lower the price, the more is the incentive to buy, and therefore the bigger is $I(s)$. The point of zero imbalance $I(s_0)=0$ is where supply and demand balance each other and therefore is the point of fair value.

In the process of repricing, function $I(s)$ experiences changes. In the process, it may evolve smoothly, or turbulently. In the smooth process the monotonicity of $I(s)$ is preserved. In the turbulent process, $I(s)$ is not always monotonous. Moreover, the imbalance function may have more than one zero: $I(s) = 0$. These are the signs of an unstable market, in which fair value is not well defined.

Smooth transition - small price shift, no formation of multiple fair values



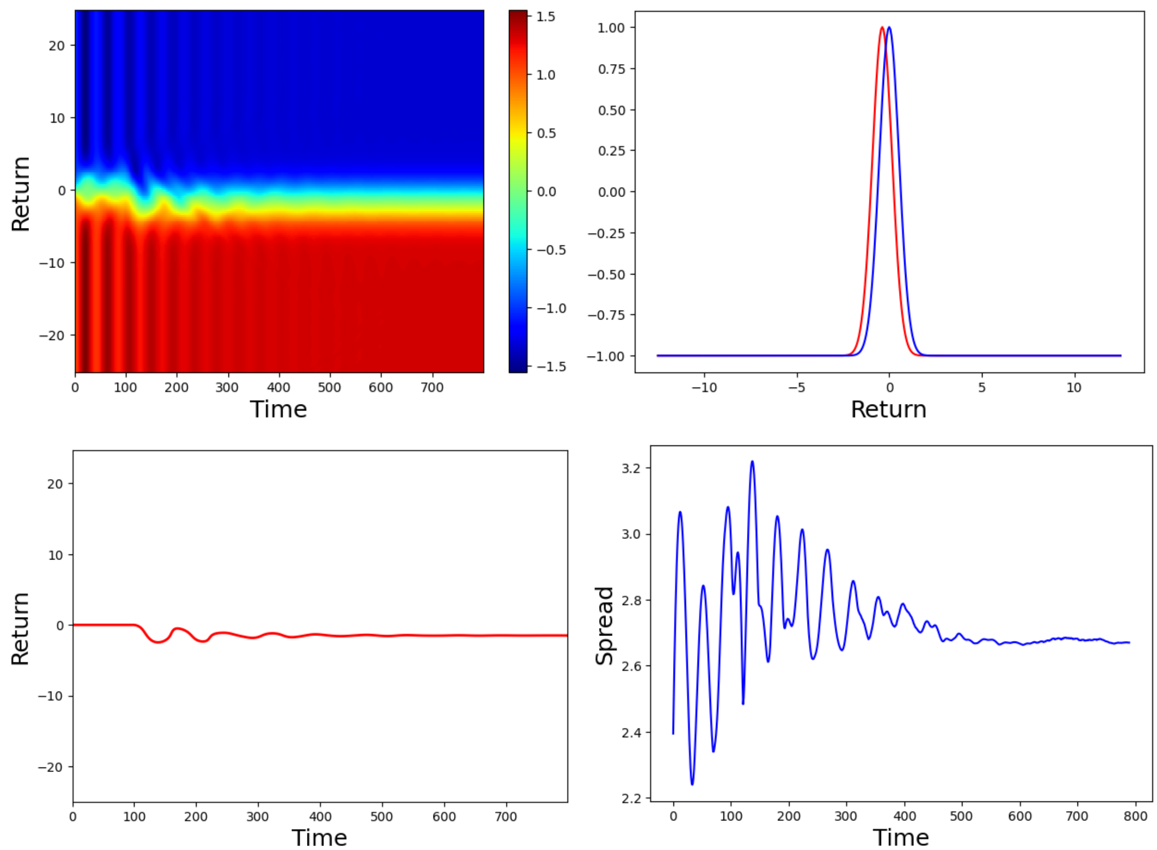
Turbulent transition - large price shift, multiple fair values emerge



The examples below show imbalance and dynamics for different price shifts. When the price shift is small (d=1.5), the transition to the new steady state occurs smoothly. Such small repricing does not inflict enough impact on the order book to create a spread spike.
The case of a moderate price shift (d=5) shows how the multiple zero-imbalance points are beginning to form, leading to price oscillations up and down. Spread has a spike that relaxes to pre-repricing levels after about 300 time steps. Lastly, the transition in case of a large price shift (d=15) is quite chaotic. Although spread shows a larger spike, is does not help to win more relaxation time.

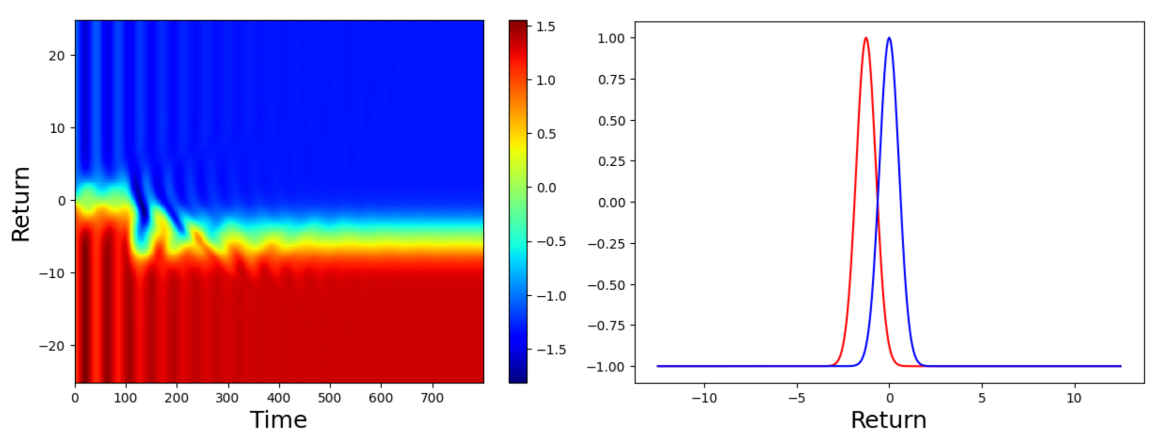
The following figures show repricing dynamics for the price shift of d=1.5:

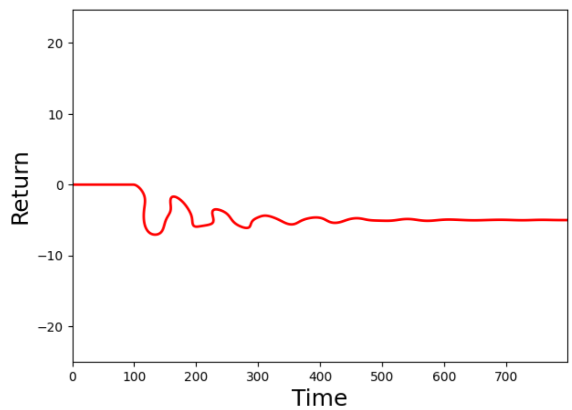
- Imbalance dynamics
- Coefficient $a_2(s)$ before and after repricing
- Zero-imbalance curve
- Spread relaxation

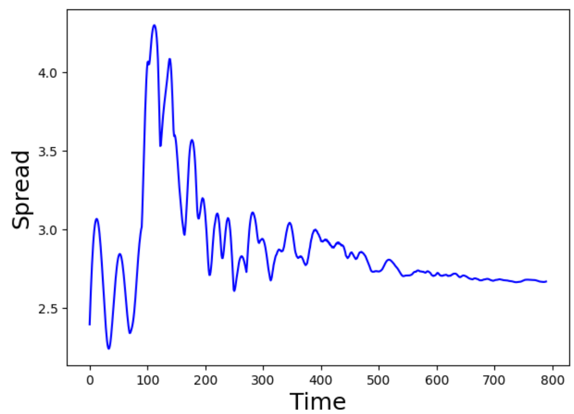


The following figures show repricing dynamics for the price shift of d=5:

- Imbalance dynamics
- Coefficient $a_2(s)$ before and after repricing
- Zero-imbalance curve
- Spread relaxation

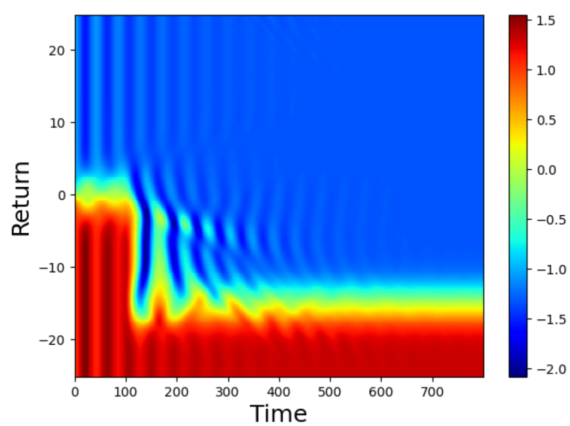


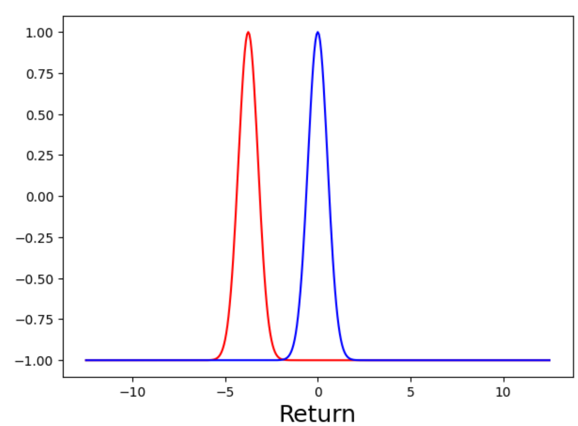


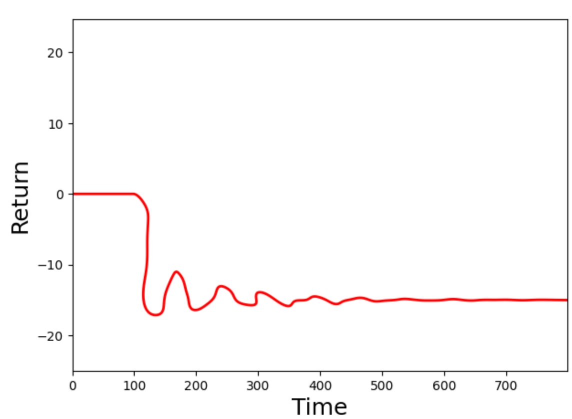


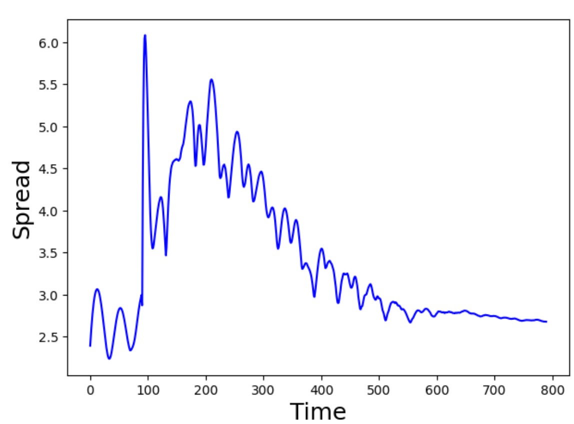
The following figures show repricing dynamics for the price shift of d=15:

- Imbalance dynamics
- Coefficient $a_2(s)$ before and after repricing
- Zero-imbalance curve
- Spread relaxation









Thus, the order size used in the destabilizing flow should be sufficient to penetrate just the characteristic depth of the order book, and not more. Smaller orders would not create a spike in the spread. Oversize orders, while creating a spike, do not give destabilizational advantage. In order to have maximum advantage, orders should be just large enough to penetrate the depth of the order book. For US equities, this usually corresponds to the top 2-3 levels of the order book.

Another question relates to order timing and frequency. In case of small orders, the relaxation time is determined by τ_2. In case of large orders, the relaxation time is determined by τ_1. For orders of the optimal size, the relaxation time τ is in between the τ_1 and τ_2. If submitted with frequency $1/\tau$, the spread will not have enough time to relax to its original value before another order is submitted, acquiring an incremental growth rate α:

$$\frac{dw}{dt} = \alpha \frac{w}{\tau_2}$$

The exact value of α is determined by the parameters of the model and can be computed. Ultimately, the spread will grow exponentially, reaching large values quickly:

$$w(t) = w_0 e^{\frac{\alpha}{\tau_2} t}$$

For example, if a normal market spread is 4 bp, relaxation time is 30 sec and spread growth rate is 3%, the spread of 1% would be achieved in 53 minutes.

RECOMMENDATIONS FOR FLASH CLASH PREVENTION MECHANISMS

1. **Coordinated Circuit breakers**
   The most standard feature functioning in the exchanges that prevents severe price moves is a circuit breaker. Circuit breaker halts trading for a period of time to allow market participants to process information, reduce the emotional component, and continue trading with better control. It is definitely one of the techniques that can prevent a drawdown from developing into a flash crash.

2. **Slowing down the markets**
   Rather than halting trading altogether, another efficient measure is to slow down the markets. This can be done by delaying order processing times by substantial amounts, for example, by 5 or 15 minutes, so that orders taken in by brokers arrive at the exchange after the delay. This will eliminate the speculative order flow and leave only the order flow that has a sufficiently large planning time horizon.
   We argue that this measure is only partial and will not eliminate the provocative order flow intended to destabilize the market. As we showed, such order flow can and must be planned, and therefore, it will pass through this measure. However, by eliminating the

speculative order flow with this measure, exchanges will automatically stop the trend following algorithms from piling up on the trend and adding their own liquidity, thus indirectly enhancing the provocative order flow.

3. **Rejecting market orders**
   During an ongoing market crash order books become sparse and thin, as orders tend to be smaller in size and spaced out farther apart from each other. By taking the already sparse liquidity in a crash, market orders contribute to further destabilization of the markets. As part of flash crash prevention mechanism, exchanges may be required to stop accepting market orders from traders.

4. **Limiting number of transactions in time intervals per account/counterparty**
   In this study we have identified trading frequency as an important factor in market destabilization. Even when they add liquidity, orders submitted regularly and frequently at the top of the order book can act against market stabilization, by causing spread to spike when they are filled and preventing it from shrinking. Therefore, it may be feasible to limit the number of accepted orders per time interval from each counterparty or account. Doing so would allow market participants to come to terms regarding the fair value of securities faster and in a smoother process.

5. **Limiting order size**
   We have also identified order size as an important parameter in provoking a flash crash. While the size of provocative orders has to be large enough, it is still limited in size. Therefore, it is feasible to reject small orders. The threshold size can be determined for each security as the average total size of its 2 or 3 top levels on the buy side of the order book. Limiting order size will also help limit frequency, which is another factor discussed in the previous paragraph.

6. **Disallow short selling**
   Lastly, enforcing a prohibition on short selling and borrowing, as it was done during the 2008 financial crisis, can also be an effective tool. During the 2008 crisis, short selling of financial securities was disallowed and proved to be an effective measure.

References

[1] Sarkissian, Jack, Statistical Mechanics of Price Stabilization and Destabilization in Financial Markets (January 6, 2020). Available at SSRN: https://ssrn.com/abstract=3514329
[2] Sarkissian, Jack, Express Measurement of Market Volatility Using Ergodicity Concept (July 20, 2016). Available at SSRN: https://ssrn.com/abstract=2812353

[3] J. Sarkissian, “Quantum Markets: Physical Theory of Market Microstructure”, Advanced Scientific Publishing (April 5, 2021)