XVI

Volume XVI, Number 2, November 2013

Journal of Applied Economics

Paul Brewer
Jaksa Cvitanic
Charles R. Plott

Market microstructure design and flash crashes: A simulation approach

UCEMA

Edited by the Universidad del CEMA
Print ISSN 1514-0326
Online ISSN 1667-6726

Journal of Applied Economics. Vol XVI, No. 2 (November 2013), 223-250

MARKET MICROSTRUCTURE DESIGN AND FLASH CRASHES: A SIMULATION APPROACH

Paul Brewer
Economic and Financial Technology Consulting

Jaksa Cvitanic* **and Charles R. Plott**
California Institute of Technology

Submitted December 2012; accepted September 2013

We study consequences of regulatory interventions in limit order markets that aim at stabilizing the market after an occurrence of a "flash crash". We use a simulation platform that creates random arrivals of trade orders, that allows us to analyze subtle theoretical features of liquidity and price variability under various market structures. The simulations are performed under continuous double-auction microstructure, and under alternatives, including imposing minimum resting times, shutting off trading for a period of time, and switching to call auction mechanisms. We find that the latter is the most effective in restoring the liquidity of the book and recovery of the price level. However, one has to be cautious about possible consequences of the intervention on the traders' strategies, including an undesirable slowdown of a convergence to a new equilibrium after a change in fundamentals.

JEL classification codes: G17, G18
Key words: market microstructure, flash crash, high frequency trading, call markets, market regulation, market simulation

---

* Jaksa Cvitanic (corresponding author): California Institute of Technology, 1200 East California Blvd., Pasadena, California 91125, USA; email cvitanic@hss.caltech.edu. Paul Brewer : Economic and Financial Technology Consulting, 4219 Shady Oak Drive, Ooltewah TN 37363, email drpaulbrewer@gmail.com. Charles R. Plott : California Institute of Technology, 1200 East California Blvd., Pasadena, California 91125 USA ;email cplott@hss.caltech.edu. The authors would like to gratefully acknowledge the support of the California Laboratory for Experimental Economics and Political Science (EEPS) and the Gordon and Betty Moore Foundation.

I. Introduction

In recent years, concern has been expressed about the effect of high-frequency trading in electronic markets. In particular, the infamous "flash crash" of May 2010 has prompted regulators all over the world to consider restrictions on trading and/or and new market microstructure designs through which trading would take place. Among the objectives is the prevention of sudden large price drops, or, at least, to make them short-lived.

This paper is focused on the consequences of alternative policies for a fixed order flow. That is, the order flow remains constant, the policies are changed and the market responses to a "flash crash" causing event are studied in the context of the alternative policies. The impact of the policies on subsequent liquidity and related volatility is thus studied within a framework in which the order flow itself does not change as a consequence of the policy or subsequent volatility. Uncoordinated order flow has a direct effect on the market and that is our focus. Clearly, order flow can react to the market or the policies themselves and those relationships can be studied, but the coordination of the order flow in response to the market could rely on an additional set of principles.

The proposed interventions we study have a purpose of increased liquidity and a smoothing effect in the times of market instability. The objectives are to mitigate price changes and volatility due to flash crashes and enhance recovery after flash crashes. Specifically, we study: (i) imposing minimum resting times in limit order books (LOB), that is, banning quick cancellations of buy/sell orders when not executed at arrival; (ii) switching to call auction markets instead of the prevalent continuous double auction markets (in a call auction all orders arriving during time intervals of specified length are collected, after which pairing of buy orders and sell orders is performed and they get executed at the price that maximizes the quantity that can be traded); and (iii) other types of "circuit breakers", that is, interventions in the market with the aim of re-establishing price stability after large moves. We use a simulation approach in which the focus is on the order flow and its interaction with the market micro structure, as opposed to strategic behavior or individual decision rules.[1] Buyers and sellers arrive randomly to the market and submit bid or ask orders in a continuous market. Orders are placed in an order book according to price and time priority and stay until executed or cancelled. Price movements and liquidity are products of the order flow. Traders' behavior that produces the flow can be interpreted as being along the lines of the competitive equilibrium model, but the appearance of traders at the market and their preferences when arriving are random, as dictated by background properties of the economy in which they are operating. For example, (the majority of) the traders could be viewed as brokers who are acting on a commission and who submit orders received from their clients. The clients, on the other hand, are motivated by a common background economy but their motivation and timing in response to that background are modeled as random. That is, for the results reported here order flow does not operate through specific, event-coordinated strategies. Thus, while our environment is a continuous market in which buy and sell orders arise from random outcomes, the environment is a flexible framework that can be readily modified, for example to have strategic traders.

We cause a flash crash by a submission of an extremely large order. We then analyze the process through which the large order influences the liquidity of the book and the instability of the transaction prices for a given structure of order flow. The method facilitates a study of the conditions under which the impact of a flash crash might be substantial and of the mechanisms that might serve to mitigate the effects.

[1] We perform our analysis using a simulation tool platform that can process, execute, and allocate in the range of 10,000 - 80,000 bids and asks per second. The tool then produces the trade price series that occurs as a result of these bids and asks, their timing, or ordering, and the trade matching rules of the marketplace. A market microstructure rule can be adapted as desired, e.g., specify that a trade occurs whenever the lowest available seller's ask price is met or exceeded by a buyer's bid price, or by contrast specify that the totality of bids and asks determine price as in a call market, or otherwise specify how the price is determined from the bid price(s) and/or the ask price(s). This flexibility is an important innovation that allows the study of different market architectures while keeping the order flow controlled as needed for market structure comparisons.
As part of the platform, there is a market simulator using the JavaScript language and a new server-side JavaScript interpreter known as NodeJS (developed by others and released as free software). Unlike other languages, JavaScript's original development as a web-browser language has lead to an asynchronous event-driven execution model derived from requirements to handle events that can fire at known or unknown times or rates. It is thought that an asynchronous model may be more appropriate for programming market mechanisms where the events are order flow related. We use a simulated time approach whereby the computer creates a time stamp for each event in a list and processes the collection of events according to marketplace rules as though the events occurred at the indicated times. This approach allows the simulation of high frequency (HF) flows without large investments or limitations due to computation times that might vary from machine to machine or from year to year. It is faster, which is important because the complexity of simulations is limited by desired waiting time for results.

While the flash crash is initiated by our deliberate action, it is the type of event that could arise on its own. For example, the continuous time nature of the market invites a randomness of the timing in which uncoordinated and decentralized traders execute orders creating an underlying randomness. That itself can be a cause of a crash, for example as many sell orders happen to appear in the market at the same time. The bottom line is, we study flash crash causes in the simplest possible framework, not necessarily corresponding to the special facts of any historical situation. The purpose is to obtain insight about what could go wrong generically.

II. Background literature

To our knowledge, there have been few academic studies of minimum resting times, or call auctions, or other measures of regulating electronic trading. Two recent papers use a simulation approach, as we do. Lee, Cheng, and Koh (2010) consider a market that consists of two types of traders, systematic traders or trend followers, and "zero-intelligence" traders. As the percentage of trend followers (who all apply a similar strategy) increases, the market prices break down and there is a decrease in liquidity. They find evidence that injecting and reducing liquidity by a market maker can both be effective. They suggest that imposing minimum resting times might be a way to control liquidity, and thus, might be helpful. However, they also find that the market maker can accumulate large losses by buying in a one-sided, falling market. Therefore, they conclude that in practice, no market maker may volunteer to participate in any such market rescue efforts unless governments are willing to underwrite some of its large potential losses. In the paper Lee, Cheng, and Koh (2011) the same authors add arbitrageurs and market makers as two additional types of traders. They claim that problems of market instability might be less about high-frequency trading per se, but rather, about the domination of market activities by trading strategies that are responding to a given set of market variables in similar ways. They offer the following conclusions:

1. Any scheme to deliberately "slow down" trading does not address the fundamental demand and supply imbalance leading to the flash crash, and it may cause more problems than it solves.
2. If there are parallel trading venues, rules to alter the speed of trading may chase away traders to other venues, and may drive liquidity out of the aggregate market. Thus, it is important for parallel trading venues to coordinate their responses to avoid creating unintended domino effects.

3. Slowing-down trading may lead to potential liquidity withdrawal due to traders' adjustments.

The classic theoretical paper Madhavan (1992) compares call auctions with continuous auctions and finds that call auctions lead to more stability and better information aggregation. This is in agreement with our simulations, in which also introducing call auctions is the most effective way of smoothing out the instability caused by a drop in liquidity. However, continuous auctions are more popular in practice, and this discrepancy between theory and practice is a puzzle that has not been fully resolved. Coppejans, Domowitz, and Madhavan (2004) find that in electronic limit order markets shocks to liquidity dissipate quickly, indicating a high degree of resiliency, which is in accordance with our results.

Equilibrium models of limit order markets (with continuous auctions) include Parlour (1998), Foucault (1999), Biais, Martimor, and Rochet (2000), Parlour and Seppi (2003), Foucault, Kadan and Kandel (2005), Goettler, Parlour, and Rajan (2005), Back and Baruch (2007), Biais and Weill (2009), and Bias, Foucault and Moinas (2013), among others. These papers aim to derive the equilibrium price formation process and most of them do not compare different market designs. In order to cope with the large dimensionality of the state and action spaces of limit order markets, these studies use stylized models with a lot of simplifying assumptions. In contrast to the equilibrium considerations, we study the formation of transaction prices given the distribution of orders, which we take to be exogenous. Dynamic, expected utility maximization models include those of Avellaneda and Stoikov (2008), Cont, Stoikov, and Talreja (2008), Kuhn and Stroh (2009), and Rosu (2009). Those studies assume specific functional forms that govern the traders' preferences. Our approach is more pragmatic: our simulation method works for any possible distribution of orders — equilibrium or otherwise.

In the theoretical literature on market microstructure, our paper is most closely related to the above mentioned Biais and Weill (2009), and Bias, Foucault and Moinas (2013). The former paper studies how, in equilibrium, limit order markets absorb transient liquidity shocks when traders behave strategically. Our paper shows, by simulation, how such shocks are absorbed in a market with myopic traders who immediately execute orders that arrive randomly (for example, those could be traders working on commission). In both papers, the traders make contact with the market at random (Poisson) arrival times. Some of the conclusions are the same in both models. Both the theoretical equilibrium transaction price and the simulated price drop sharply at the time of the liquidity shock, then gradually recover until they revert to its long term equilibrium level. The initial price drop and low level of trading are the immediate consequences of the liquidity shock. In the theoretical paper, the trade volume after the shock is low because of the increased unwillingness to trade, while in our paper it, is low because the limit order book has been depleted. That is, just after liquidity shock, the bid-ask spread is large, but, with time, limit orders accumulate in the order book and depth progressively builds up, resulting in a decrease in the spread.

In the paper Bias, Foucault, and Moinas (2013), a liquidity shock happens endogenously, because of the presence of two types of traders, algorithmic traders and human traders. When the humans' anticipation changes in an abrupt manner, the algo traders may need some time to modify the codes and parameters of their trading algorithms. The model indicates there can be a period of miscoordination during which algorithms submit orders which trigger excessive price changes. However, later the prices revert to their normal levels, as they do in our simulations.

The comparison with the above microstructure models shows that some of the implications are robust to the model choice: some of the main consequences of a crash or a liquidity shock remain the same regardless of whether the traders place orders strategically or randomly.

Of particular relevance to our study, is the science that evolved from a long history of the use of laboratory experimental methods to study the principles that govern market behavior, including price discovery, efficiency, and volatility. Using financial incentives to create markets with controlled parameters, economists have demonstrated that the underlying price discovery process is governed by the law of supply and demand. The original discovery was fundamental (Vernon Smith was awarded a Nobel Prize in Economics) and has been extended to a wide range of economic conditions, parameters, and market institutions (see Plott and Smith 2005). For example, it is well established that the CAPM follows those fundamental principles, see Bossaerts, Plott, and Zame, (2007). The classical settings of experimental markets were generalized by Alton and Plott (2007, 2008), henceforth AP (2007, 2008), to include the study of markets in which the arrival of traders in the market is stochastic. Our model is based on AP (2007, 2008). The basic supply and demand continuously change according to the randomness of traders' private values.

III. Economic environment

Our study is based on the model of AP (2007, 2008), in which buyers and sellers arrive randomly to the market for units of one asset, according to independent Poisson processes with given arrival rates. Each buyer/seller is assigned a random reservation value for the trading asset. For example, a buyer with assigned value "x" would not pay more than "x" for one unit, but is willing to pay less. A possible interpretation is that she can sell the asset outside of the market at her reservation value (which was, in fact, the case in the experiments performed by AP 2007, 2008). The reservation values are drawn randomly from fixed distributions. In the simulations, we draw the reservation values in the iid (independent and identically distributed) fashion mostly from uniform distribution, but also sometimes from normal distribution, and, in some simulations, relative to the previously traded price, conditionally on being profitable to the trader.

In the benchmark case for our simulations, the traders submit their reservation values as limit orders, with no expiry. In these simulations, each order can be viewed as having been tendered by a different buyer or seller, so the number of traders can be viewed in terms of thousands. Even though the arrival of buyers and sellers and their reservation prices are determined at random in the AP (2007, 2008) framework, they show that the concept of Flow Competitive Equilibrium (FCE) can make rough predictions about market price behavior. FCE is defined as the price at which the expected number of buys is equal to the expected number of sells during the period of simulation, and is, thus, a form of a supply-equals-demand concept.

For the majority of the simulations, our reservation values are drawn uniformly from the range of [1,100], leading to an FCE at a price of 50, there are equal number of buyers and sellers, and the initial order flow has a rate of 100 buy orders/second plus 100 sell orders/second. However, the qualitative results we obtain below have been shown to be robust with respect to various parameter values we used. That is, the effect of flash crashes and behavior of the prices thereafter under varying microstructure assumptions are robust with respect to parameter changes. Moreover, the model scales with respect to the speed of order flows and range of order values, so a particular choice of the values for those parameters also does not matter for the nature of the results.

Our goal here is not to describe the exact time series properties of the price, but rather, the mechanisms through which the impact of flash crashes take form, the qualitative impact of flash crashes and the qualitative comparison of policies implemented to reduce those impacts. Nevertheless, let us mention that our conclusions are broadly consistent with a recent comprehensive study by the UK government, available at http://www.bis.gov.uk/assets/foresight/docs/computer-trading/12-1086-future-of-computer-trading-in-financial-markets-report.pdf , and based on the input of a very large number of researchers. The methods used vary between theoretical, empirical, and experimental, and it is encouraging to see some common conclusions coming out of such a variety of approaches, including our simulation approach.

Let us also add that our analysis is based on a theoretical model and the sensitivity of conclusions to the clearly stated assumptions can be tested. We tested the theory, using simulation due to theoretical complexity, for robustness to changes in functional forms of the distribution of random orders, and we find that the flash crashes and structure of the subsequent recovery persist. The fact that changes in functional forms of distribution do not affect the broad qualitative properties under study strongly suggests that calibration to beliefs about field market parameters, would not alter the conclusions. This methodology avoids the costs, effort and ambiguity associated with setting up a calibration. This is important because the knowledge of the buy/sell distributions, which we take as uniform or normal in our examples, might not be known or easily measured in any market found operating in the field.

IV. Flash crash effects

In this section, we consider the risks associated with a flash crash occurring under a specific market structure. As explained above, we perform simulations of limit order markets with traders who have private values drawn randomly from given distributions (one for buyers and one for sellers) and who arrive to the market at random times. We add to this market one large order to cause a flash crash, and study the properties of the order book after the crash.

We find that the liquidity in our limit order markets is summarized by the following principles:

1. An appropriate measure of liquidity is the depth (the size of the queue) of the order book at price points away from the prices at which trades are taking place. Orders that would otherwise move the market price are stopped at the price points where the depth is sufficiently large.

2. The size of the queues is governed by the arrival rate and the departure rate of limit orders (due to trades or cancellations).
3. Events that reduce the depth (the queue size) on either side of the book contribute to subsequent price volatility until the natural queuing process allows the liquidity to return. Such events include: (i) ``Flash Floods" — the accidental arrival at the same time of one or more very large orders with limits considerable off market; and (ii) Liquidity Erosion — asymmetric arrivals on one side of the book, including short-lived high-frequency orders, that have limits off the market, having the effect of reducing liquidity at several price points off the market. (iii) Change in market fundamentals, that is, in distributions of the traders' values for the asset.

At a general level, we offer the following conclusions:

1. A flash crash can leave the market eroded, especially if it has a negative effect on traders' beliefs.
2. The nature and impact of liquidity erosion is sensitive to the market structure.
3. The exact nature of crash and book erosion depends on the structure of the order flow.

By introducing a large market sell order, we cause an immediate crash by removing liquidity in the book. In our benchmark simulations, we observe that prices tend to recover quickly, as they do in most of variations on the model, except they do not recover if there is a shift in the market parameters (for example a decrease in the frequency of buy orders after the crash, or a change in distributions of traders' private values). After a crash, the order book contains fewer orders, which makes the market vulnerable to increased variance and further low prices. The robustness of a recovery cannot be judged from transaction prices alone, because it is the orders in the book, or lack of book orders, that create the potential for renewed weakness.

Change in fundamentals and order frequency. In the simulation for Figure 1, we start with the fundamental demands and supply orders drawn uniformly from [40,60] with the same frequencies of draws for demand and supply at 100 per second. Those fundamentals result in a FCE price of 50. A large sell order is submitted at the time = 150. As a result of this order, we assume that the fundamentals supporting supply and demand shift immediately from [40,60] to [15,60], an increase in supply and decrease in demand. The flash crash occurs immediately. The supply order flow continues at 100 per second and the demand order flow is reduced from 100 per second to 50 per second. The market prices stick on orders that had accumulated in the book and after exhausting those, continue the fall until they are in the range of the lower FCE of 30. In Figure 2, the simulation is from the same parameter configuration without the large sell order. As can be seen, the market prices adjust immediately to the change in demand and supply fundamentals and fall until a lower support is found it the book. The market then works its way through the book and continues the downward fall to near the new FCE of 30. It is interesting to observe that the flash crash apparently simply speeds the market adjustment to the new and lower equilibrium. Basically, the flash crash contributes to the removal of old orders in the book that provided price support for a short time. In conclusion, the immediate drop in prices is due to the change in beliefs of the traders. If this change is due to an errant trade and only temporary, as may be the case in Figure 1, and not due to a real change in fundamentals, then a regulatory intervention might be helpful to restore confidence. However, if the flash-crash is a symptom of a real shift in fundamentals, as may be the case in Figure 2, then attempts to correct it will just slow down the market adjustment to the new fundamentals.

Figure 1. Shift in fundamentals and a large sell order

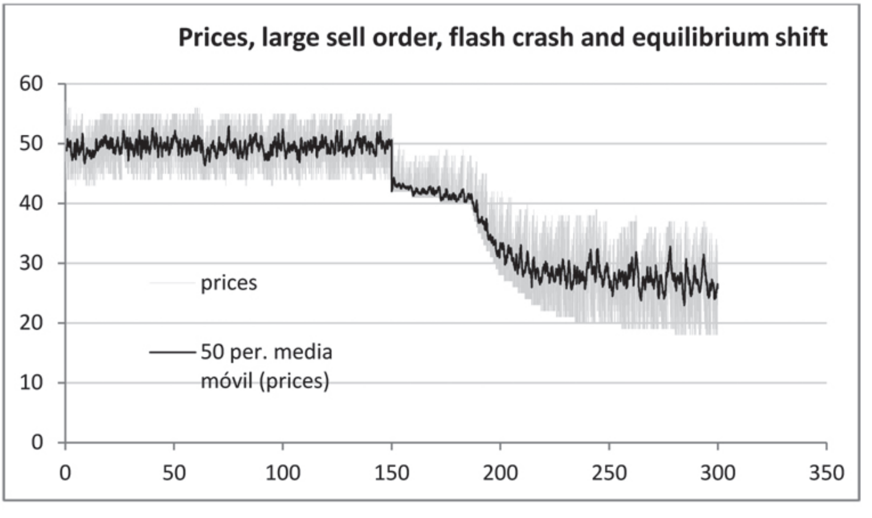
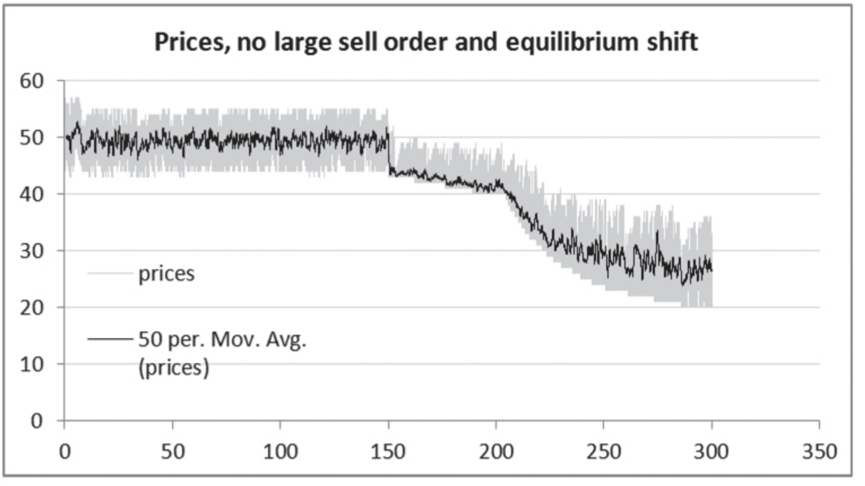
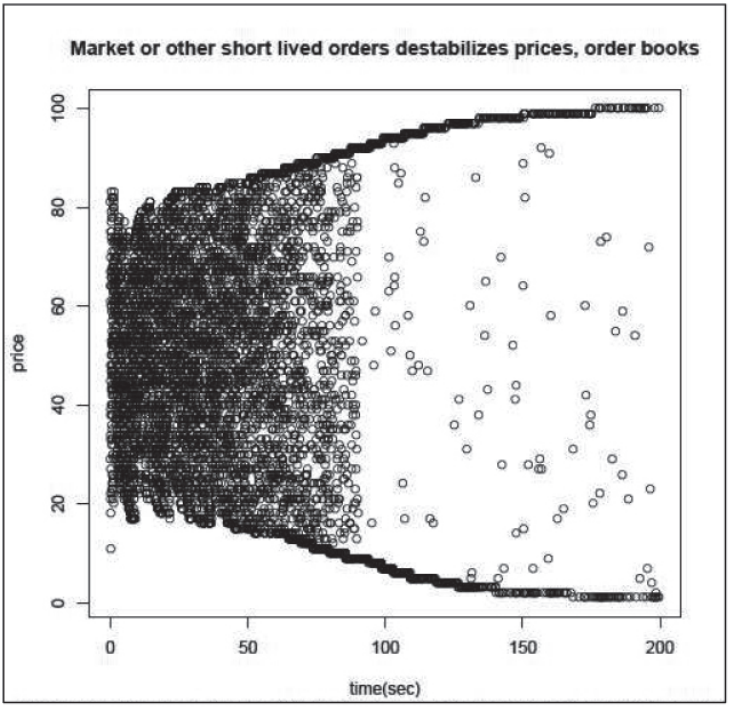


Figure 2. Shift in fundamentals and no large sell order



Short lived orders. In this simulation, each trader initially submits orders to the market as a Good-Till-Cancelled (GTC) orders (no expiry). However, each 10 seconds, 10% of buyers and sellers begin placing a 0.001 second expiration limit on their orders. This 0.001 second expiration has virtually the same effect in our model as fill-or-kill orders. The order is only available in the book for a brief moment, and usually shorter than the expected arrival of the next order. At the end of 100 seconds, all buyers and sellers are sending in orders on short expiration. As a consequence, trade slows considerably and most trades occur at the extreme prices created by leftover GTC orders. This is illustrated in the transactions prices reported in Figure 3 showing that as the transaction rate slows, prices tend to spread out towards the extreme values of 0 or 100 and often take these extreme prices, only occasionally hitting a price in the middle.

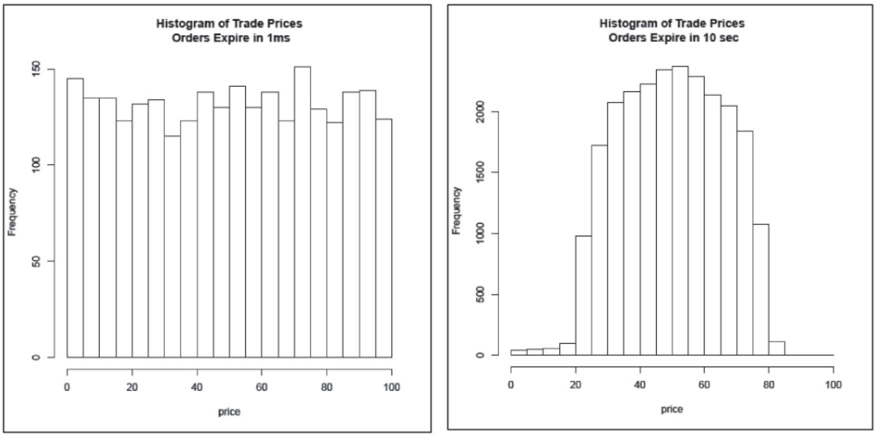
Figure 3. Reducing the percentage of no-expiry orders destabilizes the market



This happens because, when traders switch to shorter order expiry, the bid/ask spread expands as the liquidity near the equilibrium gets removed. In other words, the orders that have a longer life build up the book, and if a substantial part of the orders are changed from having a long life to a short life two things happen. The long life orders are removed from the book through trades (which would happen under any circumstance), but the short life orders that missed a trade do not remain in the book to create liquidity. If the short life orders are in an identifiable range, one can see only a small buildup of orders at the prices in that range. Outside the range one can see larger queues in the book. So, looking at a snapshot of the book one sees a price with a big queue in the book and even bigger queues as one gets further from equilibrium. However, between these two big queues there are only small queues. It looks like the "hook'em horns" (index and small fingers extended and two middle fingers not extended). The extended fingers are the size of the order book on either side of the price range of the short life orders. These big queues of orders provide the liquidity that keeps the market inside the horns. The hook'em horns phenomena with the horns that do the hooking (corral the trades) become wider as the range of the short term orders becomes wider. In short, changing the life-span of orders has a transforming effect on the market.

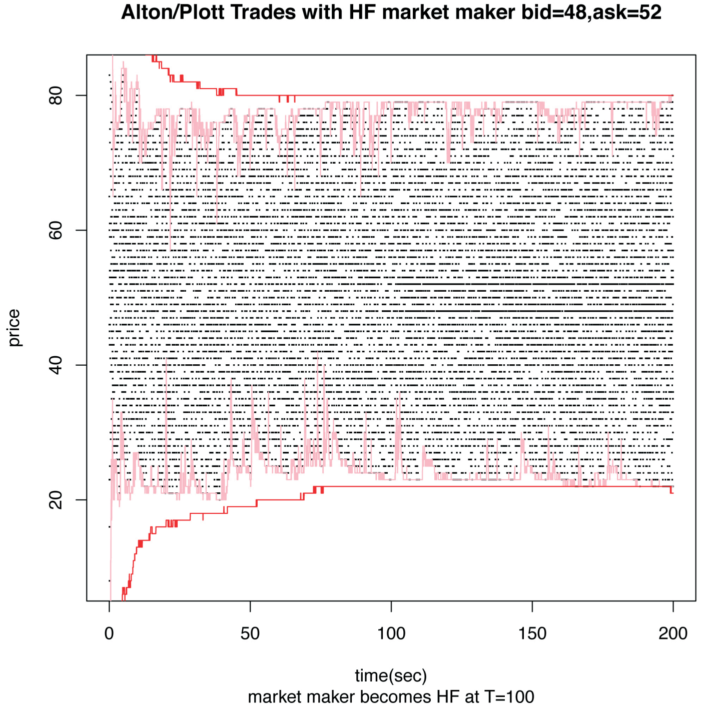
Figure 4 bins the prices in a simulation in which all buyers and sellers use the same expiration time for orders. The height of the bar shows the number of times a particular trading price was observed. The expiration time is varied from 1 millisecond (ms) to 10 seconds (sec). These all use the AP framework of traders submitting randomly price orders with reservation values iid uniform on [1,100]. The Poisson arrival rate for orders from each of the buyers and sellers is 200/seconds total. In terms of total number of trades, it appears that even fairly short expiration times can produce trading, however the trading initially spans the entire range [1,100] and is characterized as two-party trade rather than trade mediated by a market. In the figure, we see that longer expiration times lead to book formation and limit the domain of prices. Prices become less spread out, and more concentrated around the equilibrium price, and there is much more trading going on.

Figure 4. Having longer order expiration times stabilizes the transaction prices



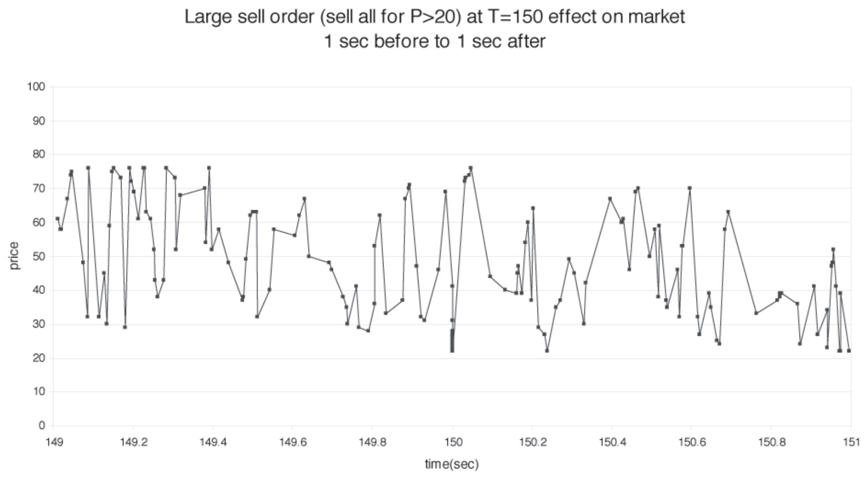
High frequency orders in narrow range. In Figure 5, we obtain different conclusions in another simulation, where we have very fast fill-or-kill, fixed-value (rather than random from an interval) orders submitted to buy at a price of 48 and sell at a price of 52. This does not have a significant effect on the price formation, other than increased number of trades around that value. This is a stylized model of HF (high-frequency) traders trying to make money by fast submissions and cancellations of specific bid and ask orders (so-called "sniping"). The observation that in this case there is no extreme effect on prices is in agreement with theoretical results of Cvitanic and Kirilenko (2010).

Figure 5. HF (high-frequency) short lived orders at fixed values do not destabilize the market



Large orders. Figure 6 shows the trading prices where a large sell order causes a flash crash, but the recovery can be fast. If the traders do not change their reservation values and otherwise behave as in AP (2007), a singular large order for 1000 units affects the prices only temporarily — the market can be surprisingly robust. The trading prices after the event are even a little higher than before the event. At 100 orders per second normal order flow, suddenly clearing the buy book of 1000 units, one would think, should take at least 10 seconds to "recover". The data show only a difficult to notice temporary effect. There are a few revisits of the low price values immediately after the event at time = 150.0 seconds, but the price data otherwise look a lot like the data preceding the event.

Figure 6. A large sell limit order at 20 submitted at T = 150 seconds



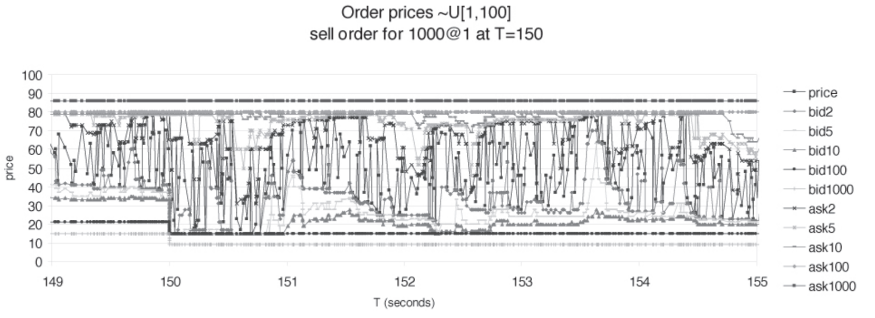
In addition to the price, we also consider **the age of paired orders traded**. Changes in the range of price movements are associated with increased age of paired orders traded. More precisely, when the change in model parameters results in orders going deeper into the order book to find trading partners, as orders that exist deep in the order book have typically been there longer, the age of paired orders increases. Thus, the age of paired orders is an indication of how long orders have been in the book before the price reached them, and is a measure of what is happening in the market. If the age is getting older, it means that structural changes are taking place even though they might not be easily visible in price patterns. In a mature market, the orders near the equilibrium price trade with each other. They are not old orders. The old orders build up in the book away from the equilibrium and provide liquidity for orders that are away from the equilibrium. This liquidity is also a cushion that keeps prices near the equilibrium and keeps variance low. A movement of trading old orders in the book reflects a process of removing the liquidity (price stability) that the accumulated orders in the book provide. Figure 7 shows the order age data, with a big spike for the large sell order event at T = 150. A few additional old orders are matched about a second later and then a few more old orders from either side of the market are matched later.

Figure 7. State of the market as represented by age of matched orders

Large market order: a "hammer". In a variation on the above, the flash crash occurs also due to a very large sell at market order, that we call a "hammer". Subsequent recovery can be measured by the shape of the market "jaws" (the number of orders in the book at various price points). Figure 8 shows the buy and sell order books just before and immediately after the simulated crash. For instance, bid2 is the 2nd best or 2nd highest bid price, bid5 is the 5th highest bid price, and bid100 is the 100th highest bid price. Similarly, ask2 would be the 2nd best or 2nd lowest asking price. In our simulation, the hammer hits at 150 seconds, at which point Figure 8 shows that the lower jaw of bid prices drops and the upper jaw of ask prices begins to jut out. This takes place because the big sell order removes the orders from the buy book leaving little or no liquidity on the down side. The liquidity on both sides was accumulated during the market maturing process. Once removed, it takes time to build up again. During that build up time, the market will exhibit increased downside variability. Low prices normally occur only at the very beginning of trade when there are few buy orders in the book. Only after the crash at T = 150 do low prices reappear, and mostly in the first second or so after the crash, although a few occur later.

We also performed simulations in which we compare the effect of the hammer in the case in which the orders come from uniform distribution to that in which they come from normal distribution. There is not much difference, because the normally distributed orders can be thought of as a one-to-one transformation of uniform orders (via the inverse normal distribution function). Moreover, in another simulation we have traders receive a value signal uniformly in [0,100] and bid or ask submitted to the market will be the last price plus a draw from a standard normal distribution, conditionally on the resulting price being profitable compared to the value signal. The crash is still self-correcting, and otherwise not too different qualitatively from the cases above.

Figure 8. Large sell market order at T = 150 seconds



We conclude that, in general, there are underlying processes of recovery that can be seen to be somewhat independent of the order generation process. After a crash, the buy book contains fewer orders. Prices from random trading can continue to touch the lows generated by the crash, until the book rebuilds.

V. Mechanisms to alleviate flash crash effects

We consider these mechanisms to mitigate the negative effects of flash crashes: (a) introducing minimum resting times; (b) switching to call auction market mechanism; and (c) shutting off trading for a period of time. They all help reduce instabilities in the market, but especially helpful, in our simulations, are the introduction of the call auction mechanism.

Minimum resting times. This mechanism involves requiring the traders to leave their orders in the limit order book at least for a minimum required duration, as to prevent them to cancel immediately the trades that are not executed at arrival.

Call auction mechanism. Most prevalent mechanism for trading in today's markets is a continuous (double) auction: as soon as a buy order (for example) arrives that is larger than the minimum sell order resting in the book, the trade is executed between those two orders, at the price value of the resting order. An alternative mechanism, often used at a beginning of a trading day, is a continuous call auction: all orders arriving during time intervals of specified length are collected, after which pairing of buy orders and sell orders is performed in a way that identifies a single price for each batch of orders that maximizes the quantity that can be traded at a single price. This price also maximizes the classical gains from trade or trading surplus. Procedurally, a simple call auction is carried out by priority-sorting the asks (from low price to high price), priority-sorting the bids (from high price to low price), and finding the intersection. In this case, sub-options include switching to call auction only temporarily after a flash crash, or having periodic call auctions.

Shutting off trading. This option stops the trading for a period of time. Sub-options include:

1. A one-time call market (Catch and Release): The policy is to suspend trading and collect the order flow while trading is suspended. The orders are collected in the book. When the market reopens the accumulated orders are treated as a call. Contracts are identified and executed. Unexecuted contracts remain on the books for liquidity when continuous trading opens after the call.
2. A series of temporary call markets. The policy is to replace the continuous market with call markets for a short period of time.
3. Stop trading, clear books and resume trading. The policy is similar to starting the market fresh. The market is delayed for a period of time; the orders that would have arrived are simply thrown away. The market is then reopened for trading with the books building from an empty slate.
4. Stop trading and take no orders, keep books unchanged. The policy here is to stop trading as before, take no trades and facilitate no trade execution, keep all orders on the books and then resume trading.

We now report on simulations in which the above mechanisms were applied.

Minimum resting times. Differing order expiration times influence both the vulnerability of the market to a flash crash as well as the subsequent healing and buildup of liquidity and consequent stabilization, as we have seen above. Longer order expiration times mean large buildup in the book at various price points. This buildup in the book creates liquidity that reduces price variance in the market. We thus conclude that requiring minimum resting times may be helpful in preventing instabilities in the market. However, this conclusion is based on very specific conditions under which we performed simulations, and it is not necessarily the case that quick cancellations by high frequency traders cause instabilities, as seen also in the experiment described above, in which high frequency traders constantly submit orders at the same value.

Call auction mechanism. We now study the behavior of call markets (CM) mechanism as a potential mechanism for trading after a flash crash. As we shall see, this results in lower variance and a more rapid recovery from a flash crash. This improved healing of the book with the CM occurs because more aggregation of orders before trading reduces the noisiness and range of the trading price, which further enhances the aggregation of orders away from the trading price into the books — orders which otherwise would become noisy trades and disappear from the book.

In the following simulation, a call market replaces the continuous double auction for the entirety of the simulation. The call market has orders accumulating for a fixed period of time, and then computes the market clearing price. Trades are executed for those orders matched with counterparties, and orders that do not trade are retained to be included in the next call.

In common with previous simulations, the following properties are maintained:

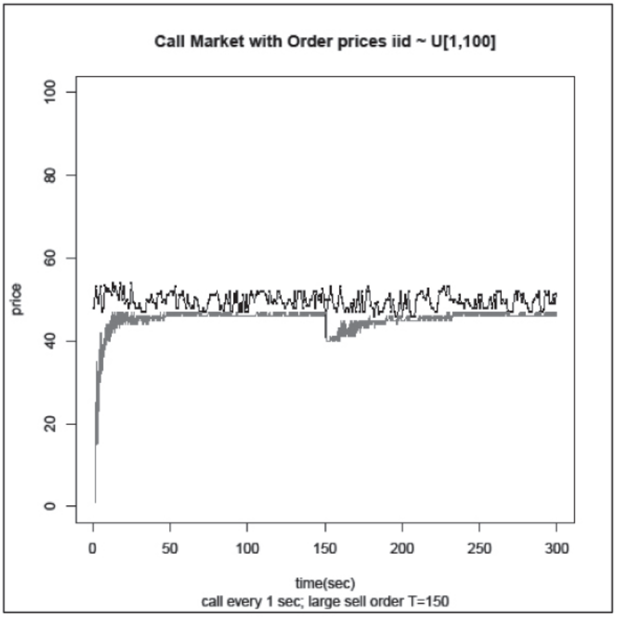
- Poisson order arrival rate of 100/second for buyers as a group, and sellers as a group;
- Orders are randomly priced iid uniform on [1,100];
- Large sell order at T = 150.

In the simulation, the call market price is determined, and the resulting trades settled, every 1 second. In Figure 9, the black line is the trade price. The red line shows the price at which 100 units are in the buy order book. Notice that from the parameters the maximum possible number of trades is 100/second (except when the large order arrives), so the black line (trade price) is always above the red price except when the large order (the hammer) arrives at T = 150.

In terms of average price behavior, the call market looks much the same as the continuous market with a book. The nature of the call is to smooth the transaction data, but in terms of a moving average the call market and the continuous market with a book appear similar.

The impact of the hammer is a fall in price, as expected. In the call market, full recovery of the “100 units” level of the buy book from the large order takes about 75 seconds, in contrast to 140 seconds for a previous simulation in an ordinary double auction. It seems there is no hope of preventing flash crashes altogether, only of reducing their impact.

Figure 9. Flash crash with call markets



Call markets could be substituted for the continuous market in the event of a flash crash. The next simulation captures the implications of such a policy. This is similar to simulations in the AP (2010) environment previously reported for the ordinary markets with books, but with the following features:

1. Initially there is no call market.
2. The list of orders studied is the same for both treatments for comparison.
3. A "Call Market Treatment" changes the market format to a call market 0.1ms after the large sell order "crashes" the market at T = 150.
4. The No Call Market Treatment, using the same exact orders as in item 3 above, does not change the market format.

As previously shown, the red lines show the price in the book at which an aggregate of 100 units are available. Figure 10 demonstrates that the implementation of the call market restores the market to near the FCE price immediately and removes the subsequent residual variance due to the crash. It is also noteworthy that the orders coming in immediately after the crash, combined with the call market, create higher prices than existed immediately before the crash. This is due to the iid uniform random nature of the order prices.

Figure 10. Transaction prices with and without introduction of call market after crash

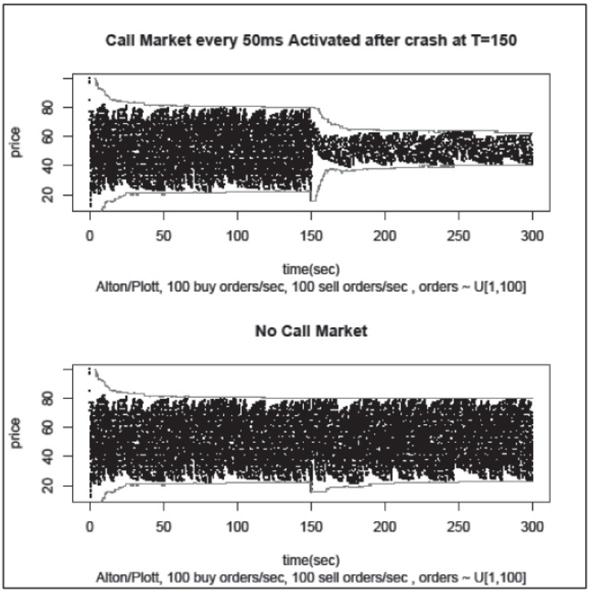
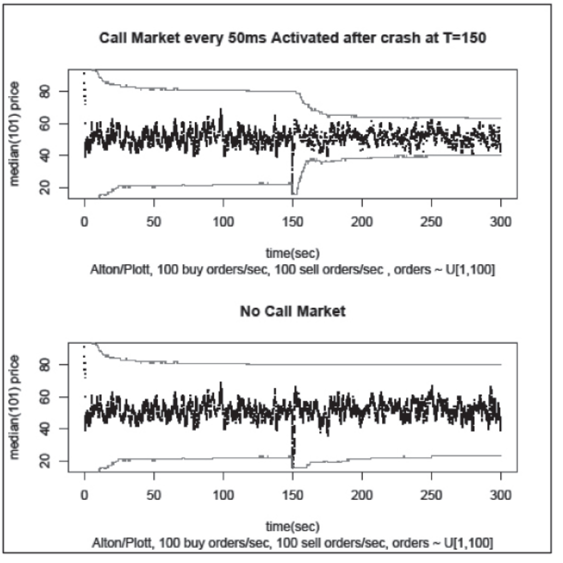


Figure 11 shows that on average, as was illustrated in the call market analysis, market price variance in terms of aggregated prices is about the same when comparing the call market with the continuous market. The median prices of Figure 11 are by definition not sensitive to the extremes of the trading prices and therefore less sensitive to certain levels of the order book that act as a bound on prices. In contrast, the extreme prices in Figure 10 are determined primarily by the health of the order books.

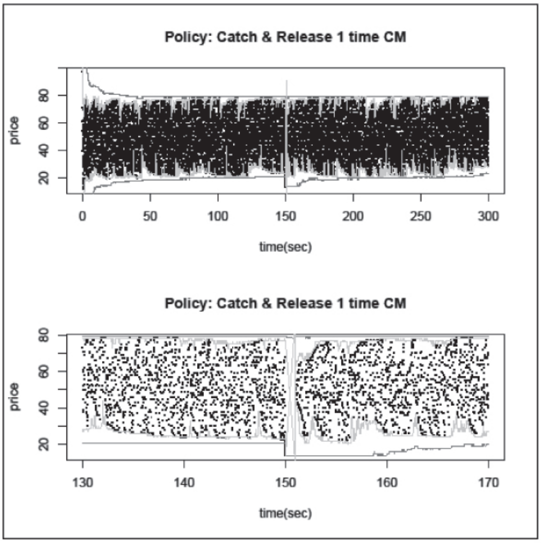
Figure 11. Moving median of 101 prices with/without introduction of call market after crash



Shutting off trading. Five different types of circuit breaker policies are examined. These can be viewed as temporary treatments as opposed to a complete and lasting change in the market structure. The same set of orders underlies all five simulations.

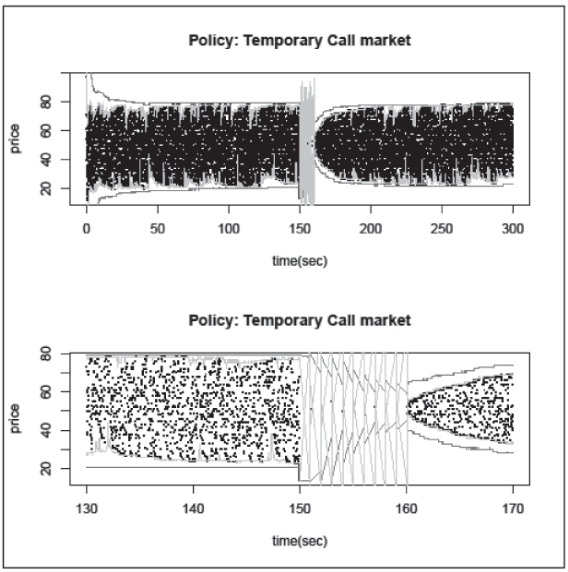
1. A one-time call market (Catch and Release). As shown in Figure 12, the liquidity removed by the hammer is replaced during the trading suspension. When the market opens, the liquidity continues to build with a consequence of subsequent small market variability. More precisely, in the catch and release policy there is substantial buildup of the book, but the variability of the continuous flow of orders has a tendency to use the liquidity far away from the market. Thus, the system experiences additional variability while the liquidity is building. The fact that the liquidity is in constant use slows its buildup. By contrast, the call protects the liquidity far from the market by coordinating and limiting trading to those orders close to the market. Thus, liquidity has a cushion near the natural equilibrium price when the call process stops. The cushion is removed by the variability once the market returns.

Figure 12. One-time call market (CM): "catch and release"



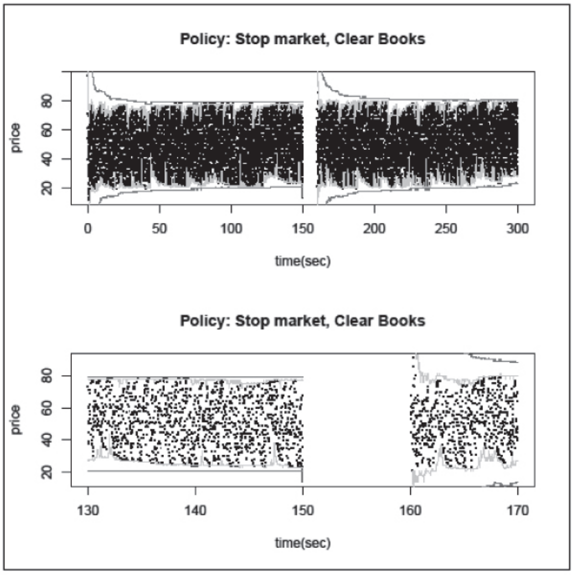
2. A series of temporary call markets. The liquidity is naturally replaced in the call market because the orders far away from the market accumulate rather than execute. Figure 13 shows the accumulation of liquidity and the subsequent return of the market to before crash conditions and behavior.

Figure 13. Temporary call market then standard trading resumes



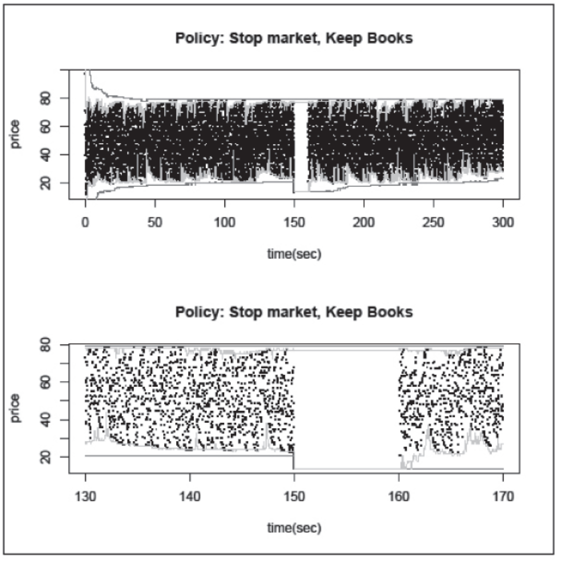
3. Stop trading, clear books, and resume trading. In the cases “Clear” and “Delay”, the market is delayed 10 seconds, the orders that would have arrived from T = 150 to T = 160 seconds are simply thrown away. The market is then reopened for trading with the books building from an empty slate. The difference is that with “Clear” the asymmetry created in the books, that can create subsequent market variability, is removed. The market buildup of liquidity is balanced because nothing is on either side of the market in the books.

Figure 14. Waiting 10 seconds with no orders or trades, then clearing books and resuming



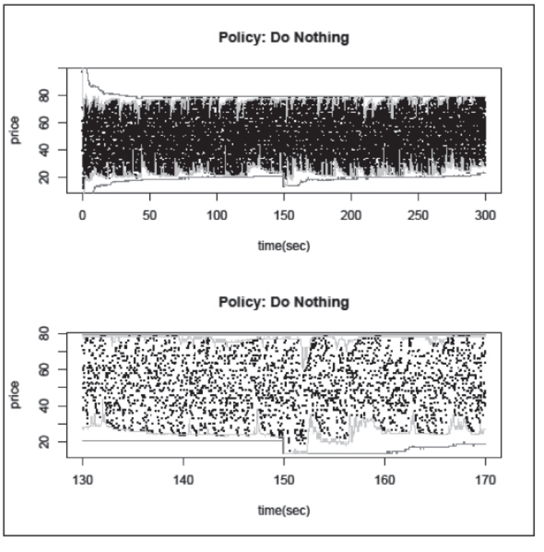
4. Stop trading (10 seconds) and take no orders, keep books unchanged. The policy here is to stop trading (10 seconds) as before, take no trades and facilitate no trade execution, keep all orders on the books and then resume trading. This maintains existing liquidity, but the asymmetry of liquidity created by the flash crash remains. As observed earlier, it requires a minute or two for the balance of liquidity to be restored and the extra price variability caused by the hammer to be restored.

Figure 15. Waiting 10 seconds with no orders or trades, keeping books and resuming



5. We also provide Figure 16 for the market where nothing is done.

Figure 16. No treatment action performed



VI. Summary

We study simulated limit order markets with buy and sell orders with random values, arriving at random times. A “flash crash” is introduced by means of a very large order. The study is focused on the subsequent effects on the liquidity of the book and the variability of the transaction prices. The simulations are performed under continuous double-auction market microstructure, and under alternative structures, including shutting off trading for a period of time, and/or switching to call auction mechanisms. While all of these help restoring the liquidity of the book and recovery of the price level, the call markets prove to be the most effective for the purpose.

The following is the list of the general conclusions we obtain, with references to the related figures.

- Flash crashes can be caused by change in fundamentals (Figures 1, 2). If this change is a temporary reaction to a fluke event, regulatory intervention can help restore confidence. If the change is long-term, a flash crash can help speed up convergence to a new equilibrium, and regulatory interventions might slow this process down.
- Flash crashes can be caused by events and practices that destroy liquidity, and they create subsequent volatility (Figures 3, 4, and 6–8).
- Liquidity-destroying practices include large market orders that execute immediately (Figures 6–9), and short lived orders, including high frequency orders (Figures 3, 4).
- The impact of short lived orders on the market depends on the proportion of traders using short lived orders (Figures 3, 4).
- Short lived orders close in price to supply/demand equilibrium have little effect (Figure 5).
- The severity of the impact of large orders on the market can be increased by the proportion of short lived orders (Figures 3, 4), and change of orders frequency in the subsequent order flow.
- When there are no reactions in subsequent order flow to a large order, recovery in price occurs quickly (Figures 6, 8,), but weakened order books allow the price to revisit low values (Figures 6, 8).
- Intervention in markets impacted by large orders should focus on rebuilding liquidity as measured by the order books, since healthy order books limit the range of subsequent prices.

In terms of policy, it may be helpful to require frequent traders to provide liquidity at all times, but requiring minimal resting times may not be very helpful, at least not for the purpose of combating flash crashes. Rather, if there is a significant chance that a sudden fall in prices will have a long-term disruptive effect, switching to call auctions until the prices stabilize may be helpful. On the other hand, in our simulations, most of the time the market stabilizes relatively soon even without outside interventions.

References

Alton, Michael and Charles R. Plott (2007), Principles of continuous price determination in an experimental environment with flows of random arrivals and departures, Social Science Working Paper No. 1276, Pasadena, California Institute of Technology.

Alton, Michael and Charles R. Plott (2008), The dynamics of price adjustment in experimental random arrival and departure environments, Pasadena, California Institute of Technology.

Avellaneda, Marco and Sasha Stoikov (2008), High-frequency trading in a limit order book, *Quantitative Finance* 8: 217–224.

Back, Kerry and Shmuel Baruch (2007), Working orders in limit-order markets and floor exchanges, *Journal of Finance* 62: 1589–1621.

Biais, Bruno, Thierry Foucault, and Sophie Moinas (2013), Equilibrium fast trading, working paper.

Biais, Bruno, David Martimort and Jean-Charles Rochet (2000), Competing mechanisms in a common value environment, *Econometrica* 68: 799–837.

Biais, Bruno and Pierre-Olivier Weill (2009), Liquidity shocks and order book dynamics, working paper.

Bossaerts, Peter, Charles R. Plott, and William Zame (2007), Prices and allocations in financial markets: Theory, econometrics, and experiments, *Econometrica* 75: 993–1038.

Cont, Rama, Sasha Stoikov and Rishi Talreja (2008), A stochastic model for order book dynamics, working paper.

Coppejans, Mark, Ian Domowitz, and Ananth Madhavan (2004), Resiliency in an automated auction, unpublished manuscript.

Cvitanic, Jaksa and Andrei Kirilenko (2010), High-frequency trading and asset prices, working paper.

Foucault, Thierry (1999), Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2: 99–134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel (2005), Limit order book as a market for liquidity, *Review of Financial Studies* 18: 1171–1217.

Goettler, Ronald, Christine Parlour, and Uday Rajan (2005), Equilibrium in a dynamic limit order market, *Journal of Finance* 60: 2149–2192.

Kuhn, Christoph and Maximilian Stroh (2009), Optimal portfolios of a small investor in a limit order market — a shadow price approach, working paper.

Lee, Bernard, Shih-Fen Cheng, and Annie Koh (2010), An analysis of extreme price shocks and illiquidity among systematic trend followers, *Review of Futures Markets* 18: 385–419.

Lee, Bernard, Shih-Fen Cheng, and Annie Koh (2011), Liquidity withdrawal and the "flash crash" on May 6, 2010, working paper.

Madhavan, Ananth N. (1992), Trading mechanisms in security markets, *Journal of Finance* 47: 607–641.

Parlour, Christine (1998), Price dynamics in limit order markets, *Review of Financial Studies* 11: 789–816.

Parlour, Christine, and Duane J. Seppi (2003), Liquidity-based competition for order flow, *Review of Financial Studies* 16: 301–343.

Plott, Charles R., and Vernon L. Smith (2005), editors, *Handbook of experimental economics results,* volume 1, New York, Elsevier.

Rosu, Ioanid (2009), A dynamic model of the limit order book, *Review of Financial Studies* 2: *4601–4641.*