Government
Office for
Science

Foresight

Minimum resting times vs. call markets and circuit breakers

Economic Impact Assessment EIA3

Foresight, Government Office for Science

Contents

Minimum resting times vs. call markets and circuit breakers

P. Brewer, J. Cvitanic and C. Plott

This review has been commissioned as part of the UK Government's Foresight Project, The Future of Computer Trading in Financial Markets. The views expressed do not represent the policy of any Government or organisation.

1. Objective

The objective of introducing minimum resting times for trades submitted to the market, that we address in this report, is to alleviate effects of a market ``flash crash". The hope is that requiring traders to have their bids and asks stay in the order book for a minimum amount of time, as opposed to being canceled immediately if not executed, will result in 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. The same objectives apply to other possible regulatory measures, such as using call auctions as the trading mechanism, or other ``circuit breakers", i.e., other measures to circumvent market instability resulting from order flow. Trade simulations are employed to compare alternative measures. More precisely, we simulate random arrivals of orders into the limit order book, and 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 methods facilitate a study of the conditions under which the impact of a flash crash might be substantial and the mechanisms that might serve to mitigate the effects.

2. Background

To our knowledge, there have been very few academic studies of minimum resting times, or call auctions, or other measures of regulating electronic trading. A couple of recent papers used a simulation approach, as we do. Lee, Cheng and Koh (2010) consider a market that consists of two type 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 paper Madhavan (1992) compares call auctions with continuous auctions and finds that call auctions lead to more stability and better information aggregation. 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.

In our analysis, the benchmark environment is a simulated, 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. The benchmark environment produces a base for providing qualitative answers for very generic, fundamental settings. In particular, our focus is on the relationship between institutional structure and of basic order flow in determining properties of price formation.

2.1. Background technology

We perform our analysis using one of the co-authors' (Paul Brewer) simulation tool platform that can process, execute and allocate in the range of 10,000 – 80,000 bids and asks per second. We adapted this tool to conduct the simulations reported here. The buy and sell orders in the model become bids and asks input into the market simulation tool. 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. This is a non-trivial task. A market microstructure rule could specify that a trade occurs whenever the lowest available seller's ask price is met or exceeded by a buyer's bid price, or could specify that the totality of bids and asks determined 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 led 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.

2.2. Background economic environment

Our study is based on 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 C.R. Plott and V. L. Smith, 2005). For example, it is well established that the CAPM follows those fundamental principles, see Bossaerts, Plott and Zame, (2007).

The classical studies of experimental markets were generalized by Alton and Plott (2010) to include the study of markets in which the arrival of traders in the market is stochastic. The basic supply and demand continuously changed according to the randomness of arrival x. The fundamental result of those studies is that the law of supply and demand has a counterpart as flows of orders and that price discovery is a balancing of the flows of demands and supplies. The resulting market prices are fully characterized by the stochastic structure of the underlying order flows.

Our study is based on the model of Alton and Plott (2010), henceforth AP (2010), 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. The reservation values are drawn randomly from fixed distributions. Given those values, they trade among themselves using mostly limit orders, but we can also allow market orders and orders that are followed by an immediate cancellation if not filled. The approach lends itself to the study of modifications of the market structure as well as various forms of strategic trading. For example, we can vary the arrival rates, or have different types of traders with regard to their arrival rates. In our simulations we draw the reservation values in the iid fashion mostly from uniform (flat) distribution, but also some are from normal, bell-shaped distribution, and, in some simulations, relative to the previous traded price, conditionally on being profitable to the trader.

Even though the arrival of buyers and sellers and their reservation prices are determined at random in the Alton/Plott 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. The FCE is based on an interpretation of the “Law of Supply and Demand” from basic microeconomics, i.e., that the market will behave in a manner that equates supply and demand. Knowing that the simulations are working in accord with previous research is an internal consistency check that contributes to confidence in the results of the simulations when applied to other questions, such as questions about crashes or recovery policies.

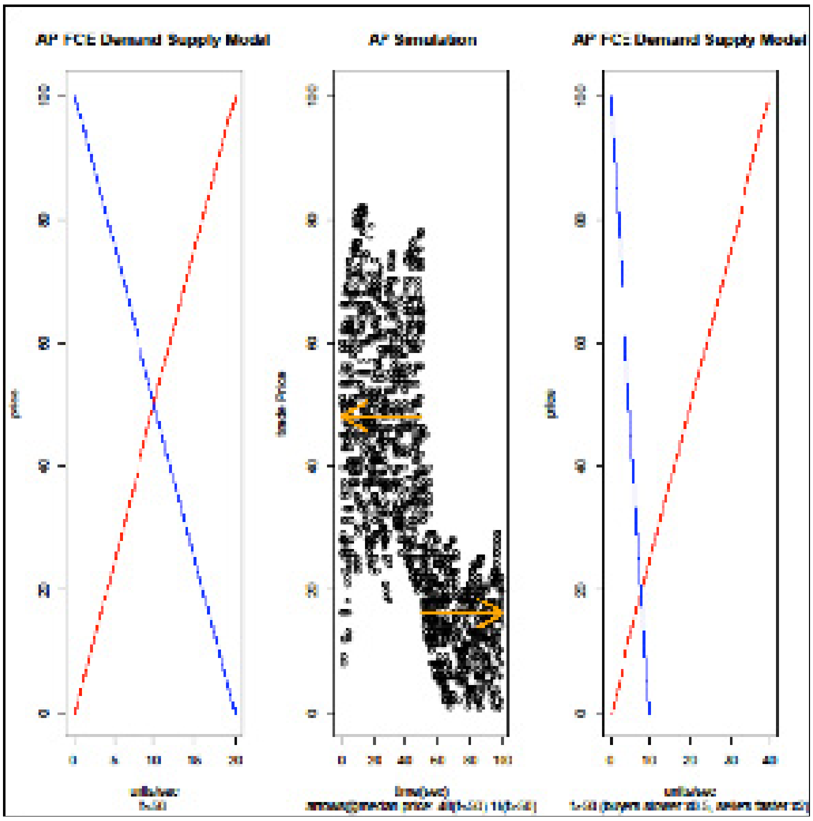
The advantage of the simulations is to demonstrate explanations of phenomena without a need for specifying various complex strategies, naive strategies, proportions of agents utilizing each strategy, beliefs of agents about state of the world, beliefs about each others’ beliefs, etc. Even if the budget, staff, and quality control were in effect to create such rich simulations reliably, actual flash crashes could also occur due to unreliable behavior of actual systems: a hedge fund with a bug in their algorithm, high-frequency (HF) traders pausing after a crash for human or other intervention (since the crash might violate assumptions the automated HF trader tests before proceeding), preferences or bugs in how market information is distributed, etc... In addition, 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 as many (say) 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, but giving insight about what could go wrong generically.

The focus of our study is on the relationship between market architecture and order flow without a need for special consideration about how the order flow itself responds to various strategies that traders might employ. An appropriate interpretation of our study is that of (the majority of) the traders being brokers, who are acting on a commission and who submit orders received from their clients. Thus we seek an understanding of the impact of the institutional and environmental framework itself on price formation before trying to understand the effects of particular trading strategies.

2.3. Benchmark case parameters

In the benchmark case 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. We use the notion of Flow Competitive Equilibrium, FCE, introduced in AP (2010), as the price at which the expected number of buys is equal to the expected number of sells during the period of simulation. For the majority of the simulations, reservation values are drawn uniformly from the range of [1,100], leading to an FCE at a price of 50. Given model parameters, prices have high variance around FCE. If we change parameters of the model, for example the frequency of arrivals, the FCE changes, and we see a shift in trading prices, as illustrated in Figure 1. The figure shows the FCE's computed from expected supply and demand at the intersections of the lines at the left and right panel, compared to the median of observed prices, marked by orange arrows in the center panel which displays the observed trading prices. The FCE model does not provide a perfect prediction, nor do we need all the technical details at this point to see that some form of rough prediction is possible. We present the FCE supply and demand and price predictions briefly in order to show that the simulations are behaving in accordance with what we know about these kinds of models.

Figure 1. Changes in order rates or distribution causes shift in trade prices



Let us finish this section by remarking that the qualitative results we obtain below, on the effect of flash crashes and behavior of the prices thereafter under varying microstructure assumptions, have shown to be robust with respect to various parameter values we used. 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 to reduce those impacts.

3. Risk assessment

In this section we explore features of the risk and the situations which can cause the risk to increase. In particular, we consider the risks associated with a flash crash occurring under the current market structure *without* the proposed measures. 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. The following is the list of the general conclusions we obtain, with information on the related figures.

- Flash crashes can be caused by events and practices that destroy liquidity, and they create subsequent volatility (Figures 2,3 and 5-14)

- Liquidity-destroying practices include large market orders that execute immediately (Figures 5-12), and short lived orders, including high frequency orders (Figures 2, 3 and 13, 14)
- The impact of short lived orders on the market depends on the proportion of traders using short lived orders (Figures 2, 3 and 13, 14)
- Short lived orders close in price to supply/demand equilibrium have little effect (Fig 4)
- The severity of the impact of large orders on the market can be increased by the proportion of short lived orders (Figures 2, 3, 13, 14), and change of orders frequency in the subsequent order flow (Figures 11, 12, 18)
- When there are no reactions in subsequent order flow to a large order, recovery in price occurs quickly (Figures 5, 7, 9, 10), and weakened order books allow the price to revisit low values (Figures 5, 7, 9)
- 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.

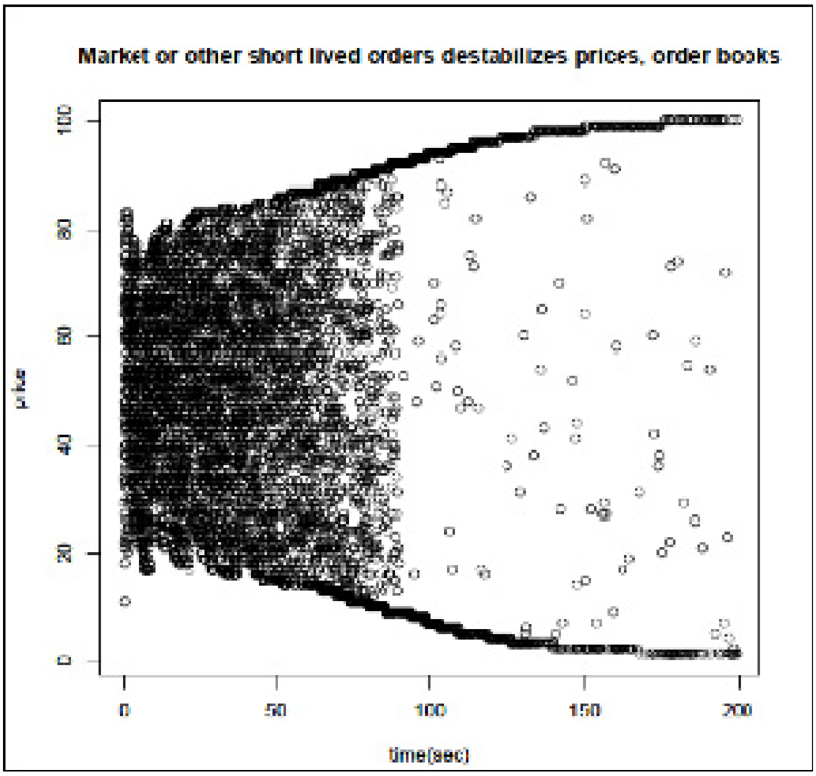
3.1. Flash crashes

The main message of this section is that flash crashes can be caused by events and practices that destroy liquidity, and they can create subsequent volatility. The examples we consider are short-lived orders and large sell orders. Short lived orders destabilize the market, if many traders use them. They do not destabilize the market if submitted in a narrow range by a small fraction of traders. Large market sell orders cause immediate crashes by eating up liquidity in the book. Prices tend to recover quickly in our benchmark model, as well as in other settings in which we vary the order generation process. 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. Finally, it should be pointed out that varying the choice of particular model parameter values, including the number of traders, does not change the nature of our qualitative results.

3.2. Short lived orders

In most of our simulations there are equal number of buyers and sellers. The initial order flow has a rate of 100 buy orders/sec plus 100 sell orders/second with values similar to the left panel of Fig 1. Each trader initially submits orders to the market as a Good-Till-Cancelled (GTC) order (no expiry). However, each 10 seconds, 10% of buyers and sellers begin placing a 0.001 sec expiration limit on their orders. This 0.001 sec 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 sec, 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 2 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.

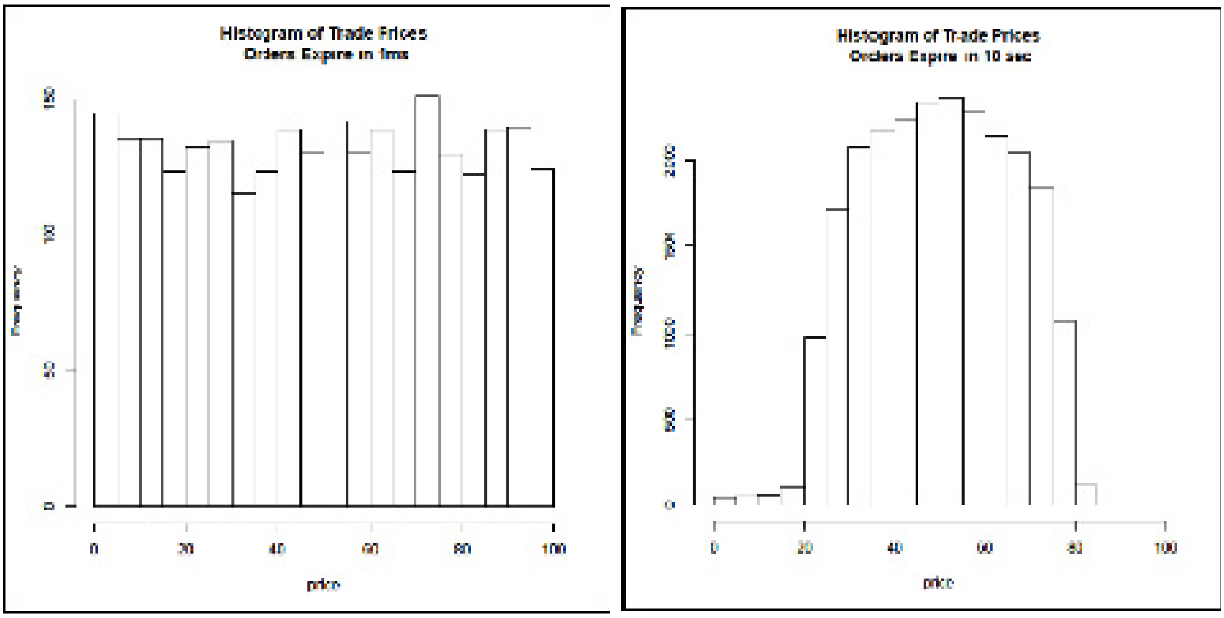
Figure 2. Reducing the percentage of no-expiry orders destabilizes the market



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, orders that have a longer life in the book 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 build up 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 the FCE. 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 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 3 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.

Figure 3. Having longer order expiration times stabilizes the transaction prices

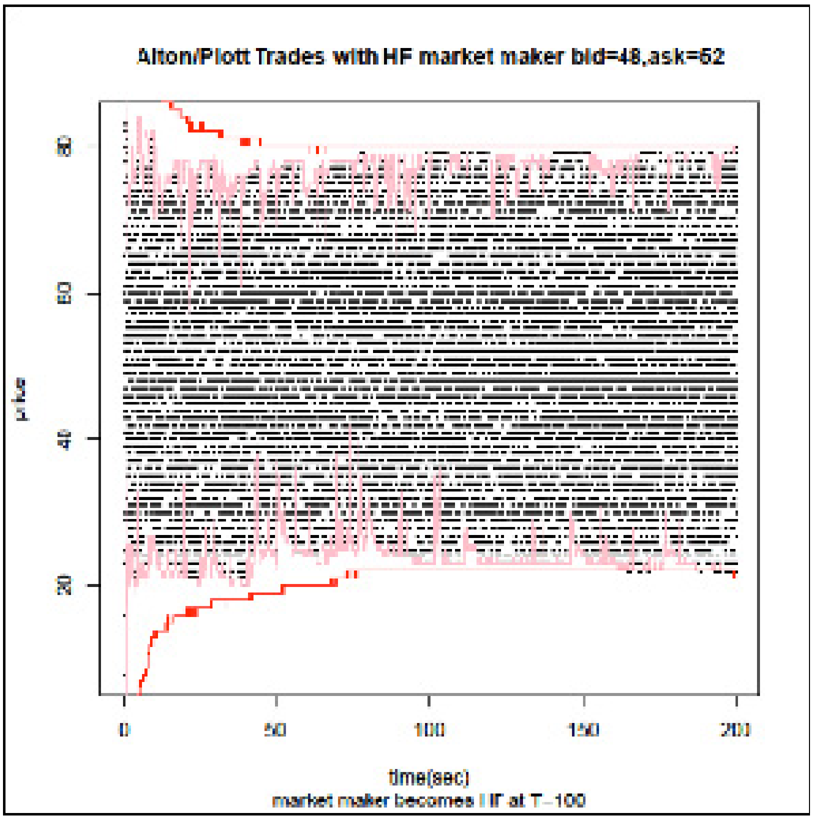


This expiration time is varied from 1ms to 10 sec. These all use the AP framework of traders submitting randomly price orders with reservation values iid uniform on [1,100]. The Poisson arrival rate for orders from each of the buyers and sellers is 200/sec total. In terms of total number of trades, it 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 figures below 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.

3.3.High frequency orders in narrow range

In Figure 4 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, and this did 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 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 from Cvitanic and Kirilenko (2010).

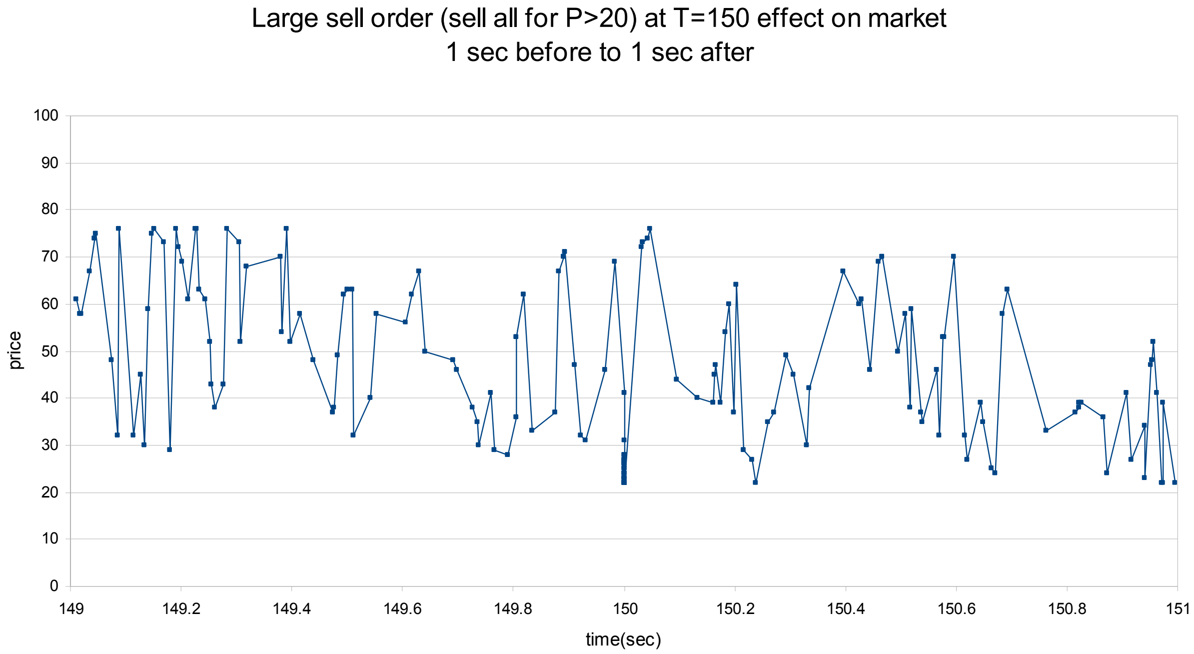
Figure 4. HF short lived orders at fixed values do not destabilize the market



3.4. Large orders

Figure 5 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 (2010), a singular large order for 1000 units affects the prices only temporarily. The following simulation shows 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 sec 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 sec, but the price data otherwise look a lot like the data preceding the event.

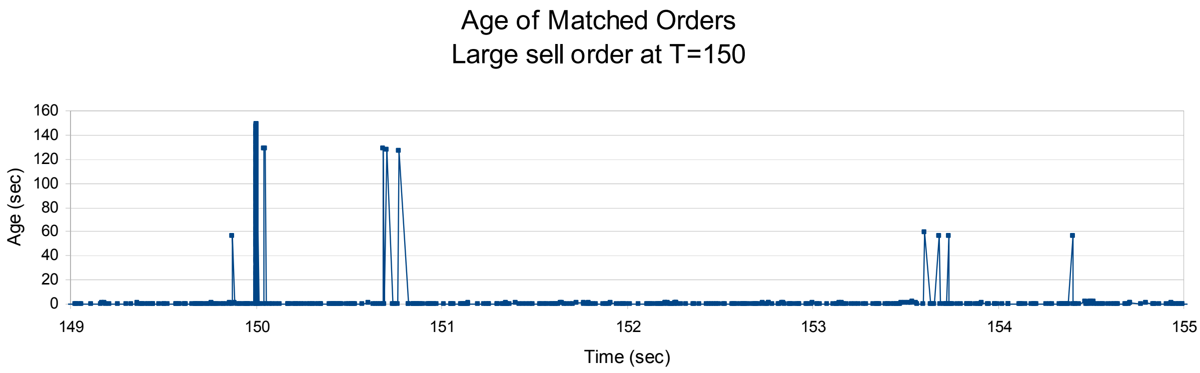
Figure 5. A large sell limit order at 20 submitted at T=150



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.

In the above described simulation, illustrated below in Figure 6, the order age data shows 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 6. State of the market as represented by age of matched orders

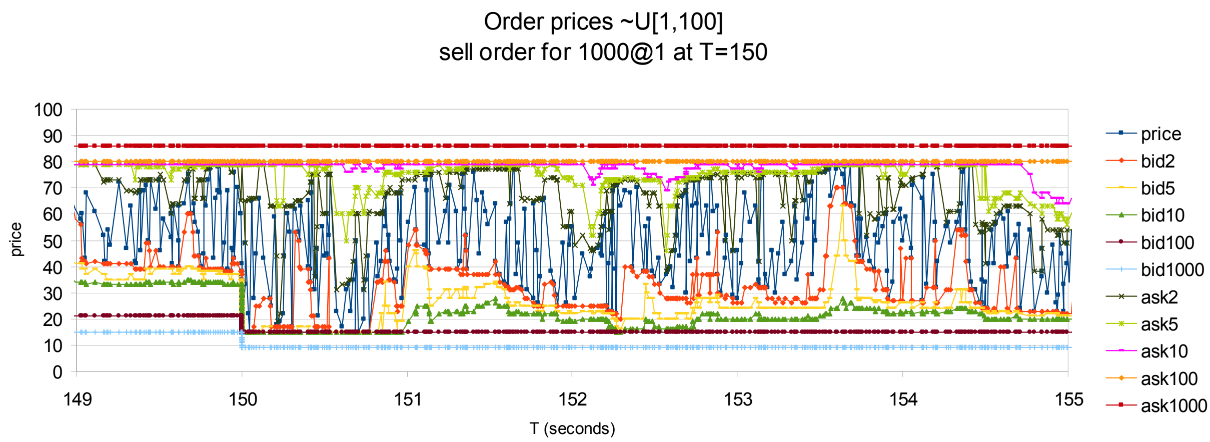


3.5. Large market order: a "hammer"

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 seen in the dynamic structure of the book and the shape of the market "jaws" (the number of orders in the book at various price points). One can imagine these orders as the excluded bids and asks in the classical demand and supply functions. The elasticity of these changes can be viewed as the changes and location of the liquidity in the market. If many orders exist at a price point the market is not going to move through that price point without some large or sustained (counter) order flow at that point. Similarly the recovery from the hammer can be seen in the buildup of the book. While this happens the market will experience downside vulnerability.

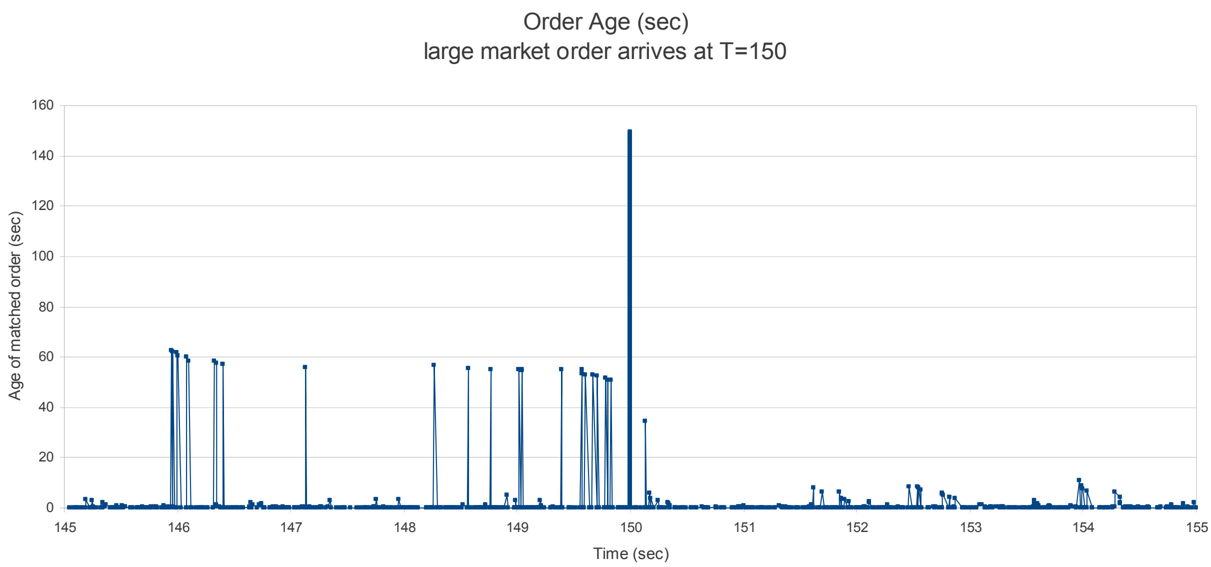
Figure 7 shows the buy and sell order books just before and immediately after the simulated crash. For instance, bid2 is the 2nd best or 2nd highest bid price, bid5 is the 5th highest bid price, and bid100 is the 100th highest bid price. Similarly ask2 would be the 2nd best or 2nd lowest asking price. In our simulation the hammer hits at 150 seconds, at which point Fig 7 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 down side variability. Low prices normally occur only at the very beginning of trade when the 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.

Figure 7. Large sell market order at T=150



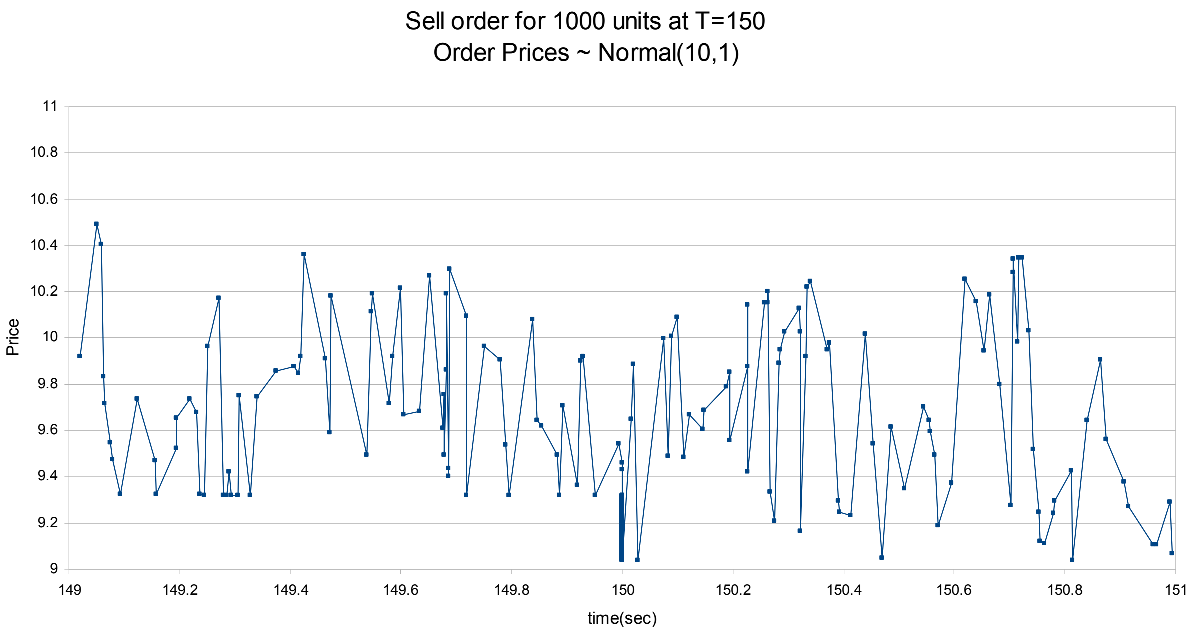
The randomness of order prices causes some eroding of the book as can be seen in the spikes of the age of matched orders caused by random sequences of sell order working their way through the lower jaw and removing accumulated liquidity. It seems as if this randomness can spill over to the shape of the upper jaw as well.

Figure 8. Age of matched orders with large sell market order at T=150



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 (bell-shaped) distribution. Figure 9 shows the sequence of trade prices: there is not much difference, which, given some thought, should not be surprising. This is because the normally distributed orders can be thought of as a one-to-one transformation of uniform orders (via the inverse normal distribution function). Thus, when matching a large market order to the book, it will reach a price p in the book measured in "uniform price units" at the same level it reaches the transformation of p measured in the "normal price units". It is not the case that one order generating distribution sets the market up for "deeper" crashes than the other when both distributions are continuous, and thus have a one-to-one relationship. For example, with the [1,100] uniform prices the 1000 unit sell order caused prices to fall to about 14-17 which is at the 0.15-0.16 p-level of the uniform distribution. So we would expect a market with normally distributed prices to fall to the 0.15-0.16 p-level of the normal distribution, which, from a lookup table, is about -1.08 to -0.95 standard deviations below the normal mean. This is illustrated in 9e note that the crash price immediately after the hammer order is about 9.05, or, since the mean of order prices is 10 and the standard deviation is 1.0, about 0.95 standard deviations below the normal mean. This is consistent with the prediction of 0.95 to 1.08 standard deviations below the mean. For the non-technical reader, this mathematical exercise can be thought of as additional confirmation of the design consistency of the simulations: that the robustness of these results from the simulated markets occurs because they are functioning in accord with known principles.

Figure 9. Transaction prices when orders are coming from normal distribution



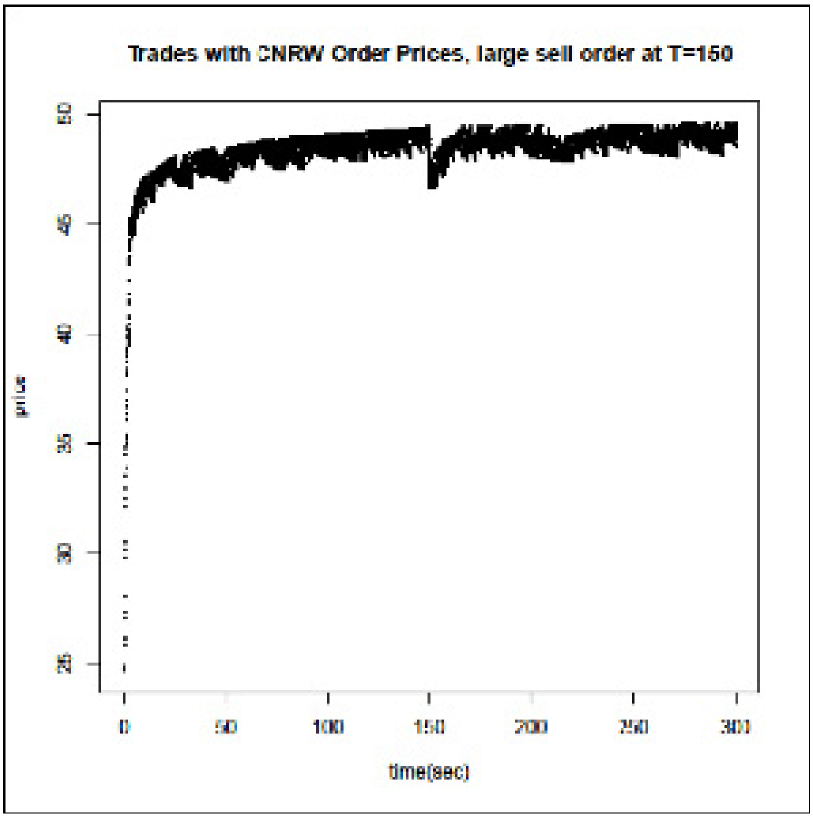
3.6. Variations of bidder strategies

The exercise illustrates that we are able to study individual bidder strategies. Reservation values are revealed to the agent who then decides what bid to submit to the market. There are two cases. In the first case traders just see private values coming uniformly from [0,1] and submit an order equal to the last price +/- 1 provided this price is profitable given their value. The market adjusts slowly with little effect of the large order because there is substantial liquidity at prices 49-51. We have not provided graphs for this case as there is not much in the way of dynamics. In a variation on this, we have traders receive a value signal uniformly in [0,100] and bid or ask submitted to the market will be the last price + randomNormal(mean 0, sdev 1), conditionally on the resulting price being profitable compared to the value signal. We call this case CNRW for conditionally normal random walk. These also need a starting price, for which we picked 25. In Figure 10, such traders equilibrate more slowly towards the FCE of 50 and display some interesting dynamics upon the crash at T=150, but the crash is still self-correcting. In the early stages of self-correcting, the prices do revisit the crash level, as is seen in the iid uniformly distributed orders and iid normally distributed orders submitted directly to the market. The prices generated by the conditionally normal bids are not as noisy as are the prices produced by the uniformly distributed orders.

We conclude that 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.

The following figure is for the CRNW case. Because of the conditional part of the pricing rule, the extra step taken by traders to base their price on the previous price, there is no prediction involving transformation of the previous results such as there was between the uniform order prices and the (non-conditional) normal order prices. However, the qualitative pattern of price change and recovery at T=150 is consistent with what happens in the other markets.

Figure 10. Prices with orders coming according to the conditionally normal rule



3.7. Big sell order with slowdown of buy flow

If a big sell order flash crash creates a slowdown in the order flow from buyers (even though the range of buy orders is unchanged) the market will experience slow recovery and vulnerability to higher variance and additional flash crashes, as illustrated in Figures 11 and 12. Before the large sell order the Poisson rate parameters of buyers and sellers are 10 orders per second. After the large sell order, the buyers' rates fall by 90% and recover on an exponential slope. The buy book will not rebuild as fast because buy orders are slowed down. Furthermore, and unlike the other cases, the relative slowdown in buy orders will allow the ask book to build an equilibrium at much lower prices than is the case without buyer slowdown. These patterns of prices occur because the flow competitive equilibrium is shifted after the crash, and then slowly shifts back.

Figure 11. Transaction prices with buy orders less frequent after the crash

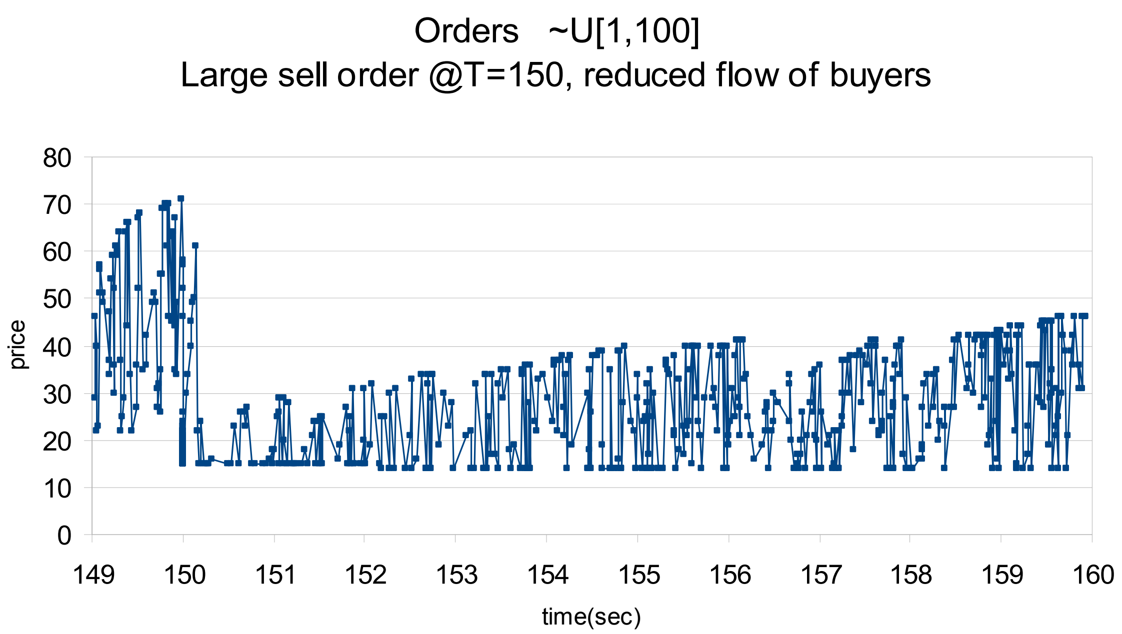
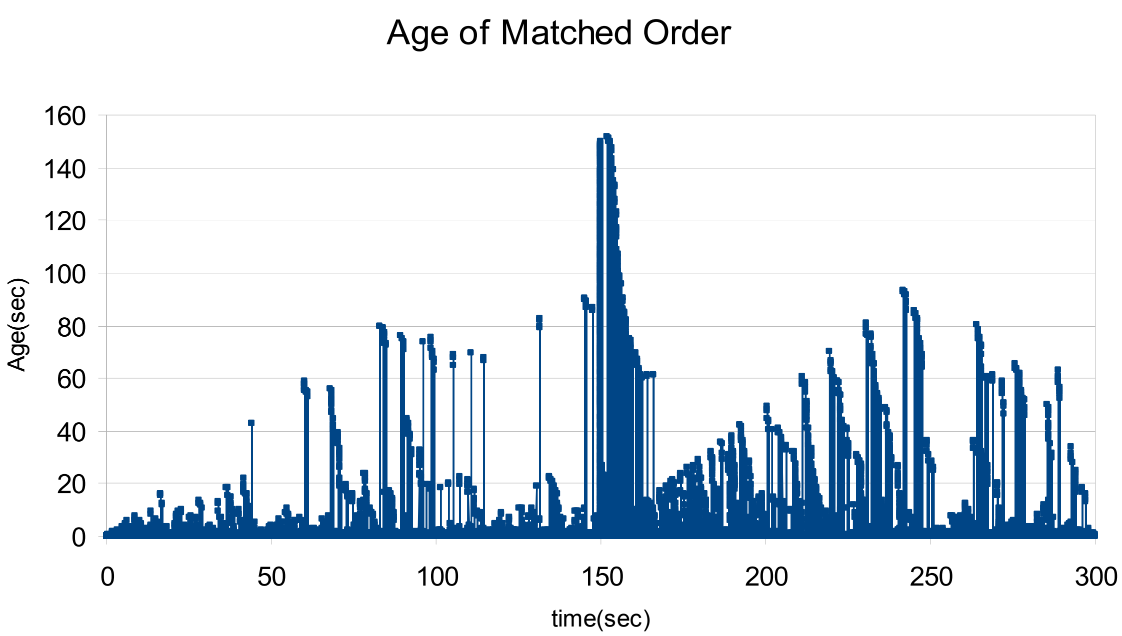


Figure 12 shows the age of orders for the scenario with a reduced flow of buyers after the crash. Recall that the crash at T=150 is typically associated with a spike in the order age because the large sell order necessarily matches old orders in the book. What is different here is the persistence of new sell orders matching old buy orders in the book, e.g. at T=160-240 and this occurs because the flow of buyers has reduced. As more buyers return to the market and place orders, the buy order book rebuilds with fresher orders.

Figure 12. Age of matched orders with buy orders less frequent after the crash



4. Options

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

4.1. Minimum resting times

This option 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.

4.2. Call auction mechanism

Most prevalent mechanism for trading in today's markets is a so-called 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 . In this case, sub-options include switching to call auction only temporarily after a flash crash, or having periodic call auctions.

4.3. Shutting off trading

This option options 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.

5. Costs, risks and benefits

5.1. 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 seen in figures 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 cancelations by high frequency traders cause instabilities, as seen also in the experiment described above, in which high frequency traders submit constantly orders at the same value.

In Figures 13 and 14 we present another pair of simulations, the first of which is our benchmark setup with a hammer, while in the second we have 50% of the traders having very short lifetimes of 0.001 seconds. In each of the figure the trade prices are shown, and the lower portion of the display zooms in on the detail around the crash at T=150. The latter corresponds to a market in which half of the traders are high-frequency traders who quickly cancel their orders, while the former can be thought of the same market with a ban on cancelations. We see evidence that the liquidity is lower, and the bid-ask spread and the price variance are larger when cancelations are allowed. There is also evidence, but less clear, that the effect of the hammer is smaller and the recovery faster when cancelations are banned.

Figure 13. Hammer effect without HF traders

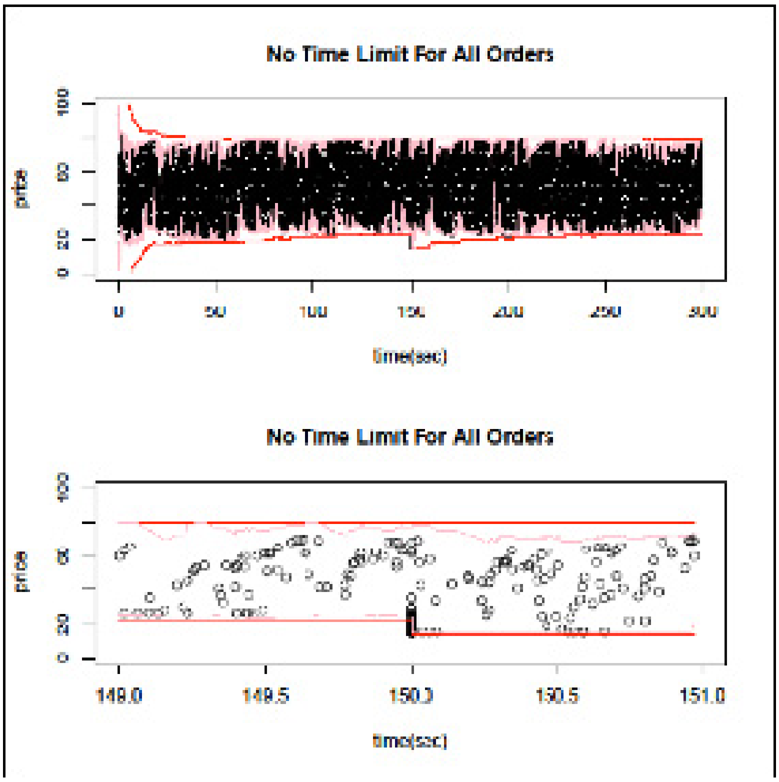
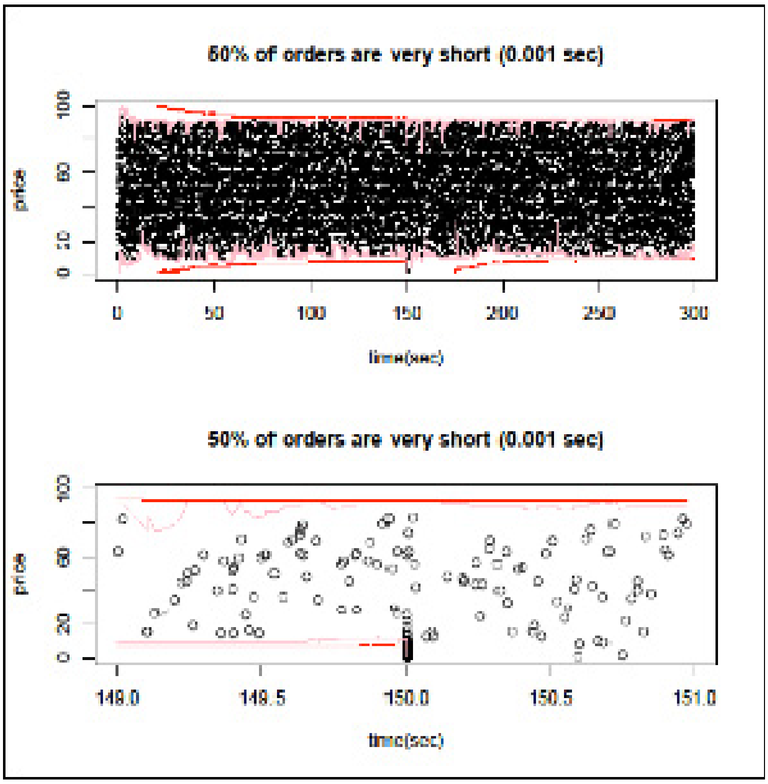


Figure 14. Hammer effect with 50% of HF traders



5.2. 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.

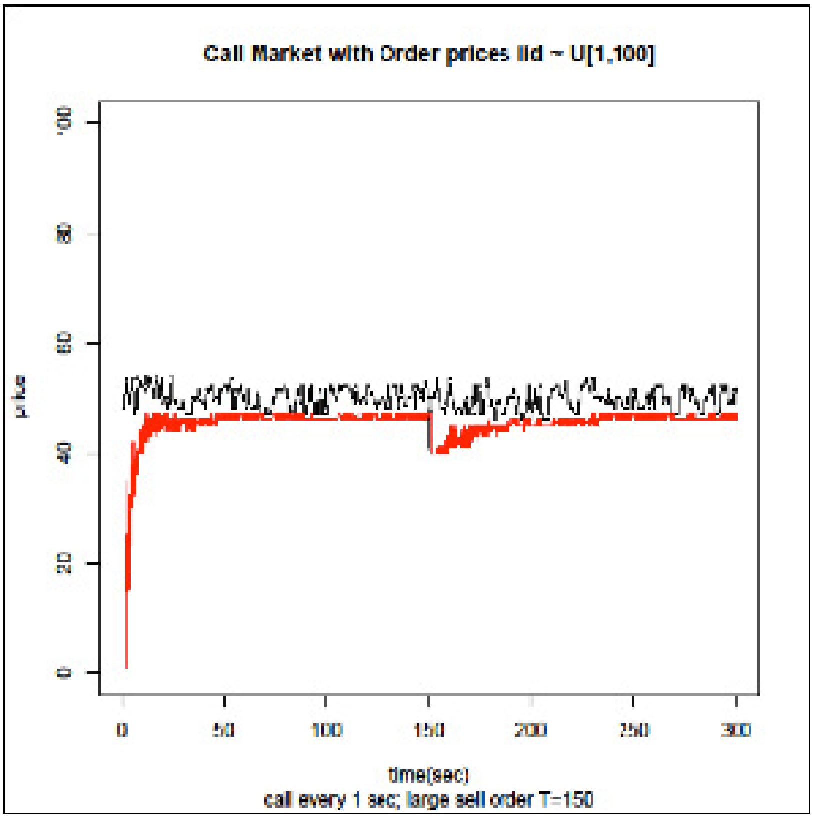
In the following simulation a call market replaces the continuous double auction for the entirety of the simulation. The call market has orders accumulating for a fixed period of time, and then computes the market clearing price. Trades are executed for those orders matched with counterparties, and orders 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/sec 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 15, 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.

Figure 15. Flash crash with call markets



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.

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 Alton/Plott environment previously reported for the ordinary markets with books 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 in item 3. above, does not change the market format.

As previously, the red lines show the price in the book at which an aggregate of 100 units are available. Figure 16 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 16. Transaction prices with and without introduction of call market after crash

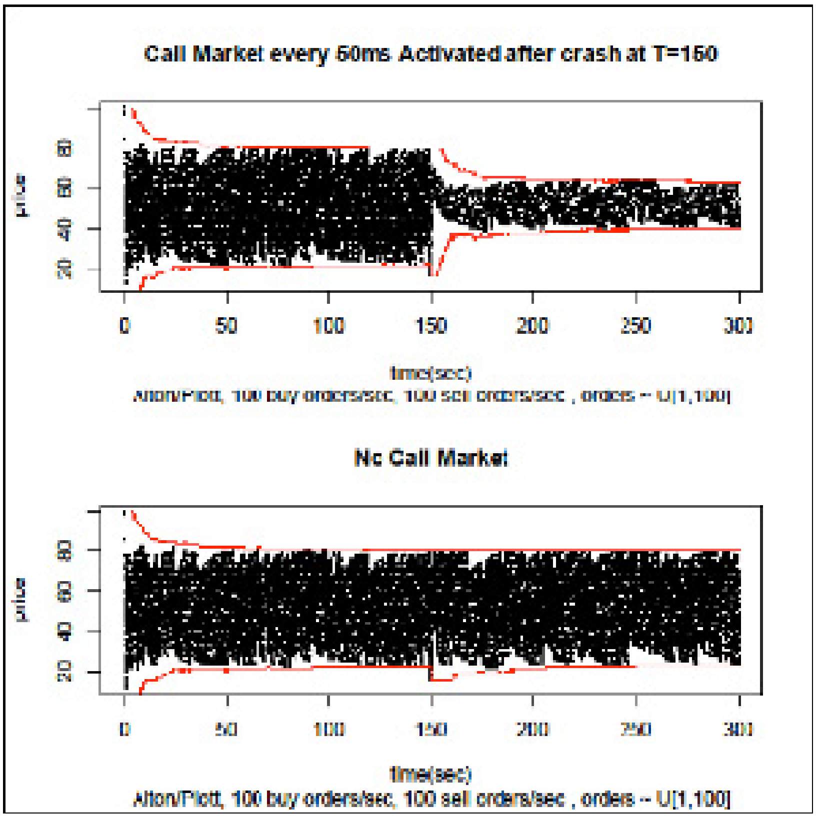
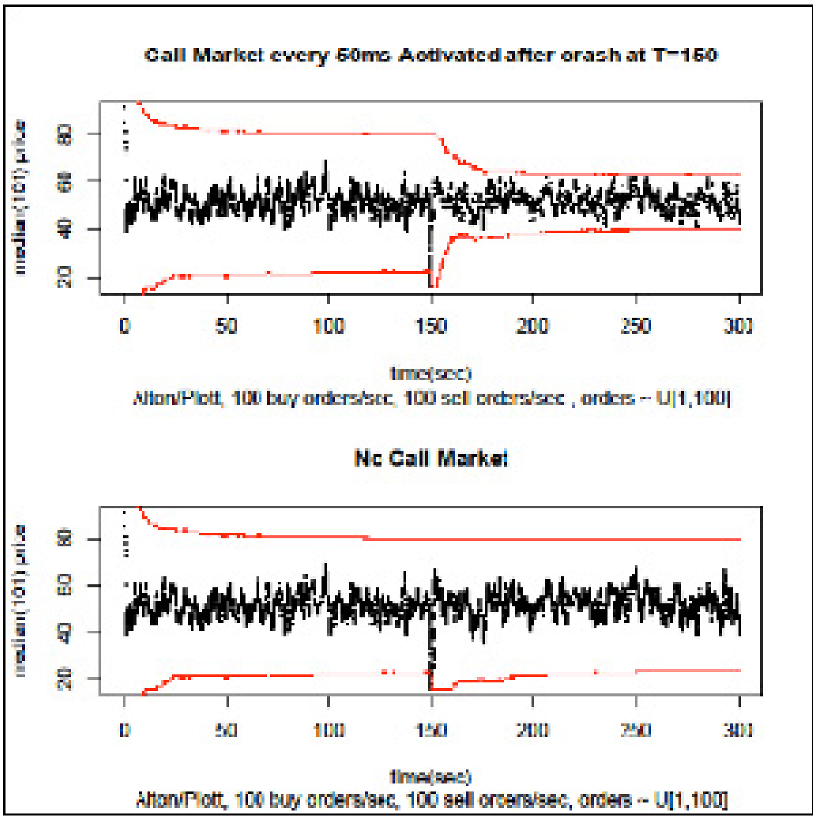


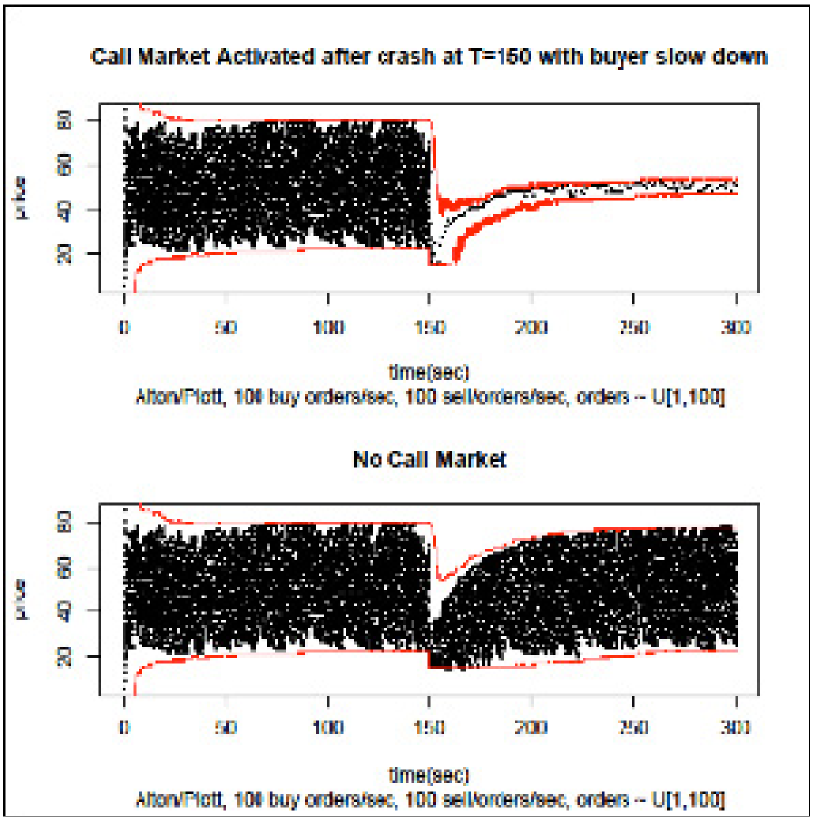
Figure 17 shows that on average, as was illustrated in the call market analysis, market price variance in terms of aggregated prices is about the same when comparing the call market with the continuous market with a book. The median prices of Figure 17 are by definition not sensitive to the extremes 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 16 are determined primarily by the health of the order books.

Figure 17. Moving median of 101 prices with/ without introduction of call market after crash



In the following simulation (Fig 18) a single set of random orders is generated having the large sell order induced crash at T=150 and the buyer Poisson rate slowing down beginning at T=150 and recovering on an exponential decay of the slowdown. Notice that the red line, which depicts the 100th unit in the buy book is flat in the top pane and shows no recovery until T~165 sec with the call market whereas in the bottom pane it is almost flat until T~190 sec without the call market.

Figure 18. Buyer slowdown with and without a call market



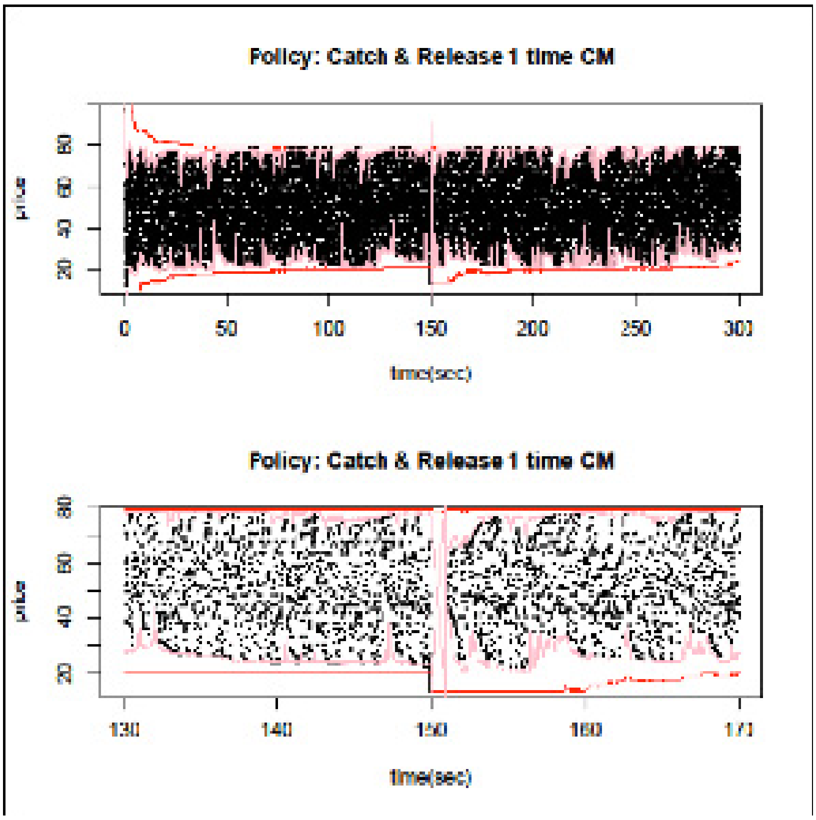
5.3. Shutting off trading

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 19, 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 build up 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 FCE price when the call process stops. The cushion is eaten away by the variability once the market returns.

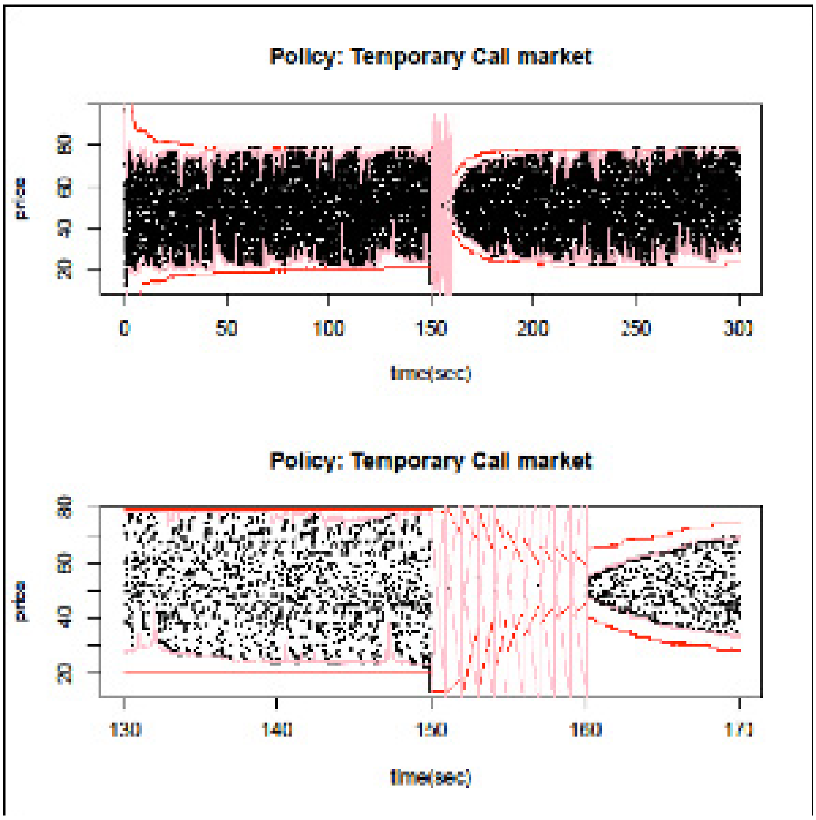
Figure 19. One-time call market ("catch and release")



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 20 shows the accumulation of liquidity and the subsequent return of the market to before crash conditions and behavior.

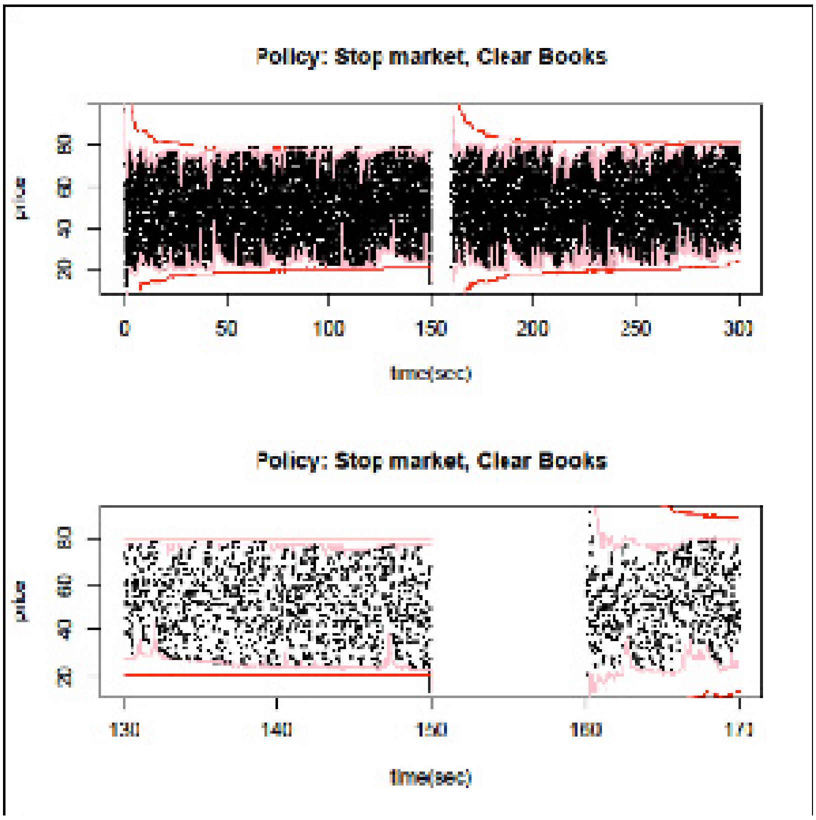
Figure 20. Temporary call market then standard trading resumes



3. Stop trading, clear books and resume trading

In the cases “Clear” and “Delay” the market is delayed 10 sec, the orders that would have arrived from T=150 to T=160 sec 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. See Figure 21.

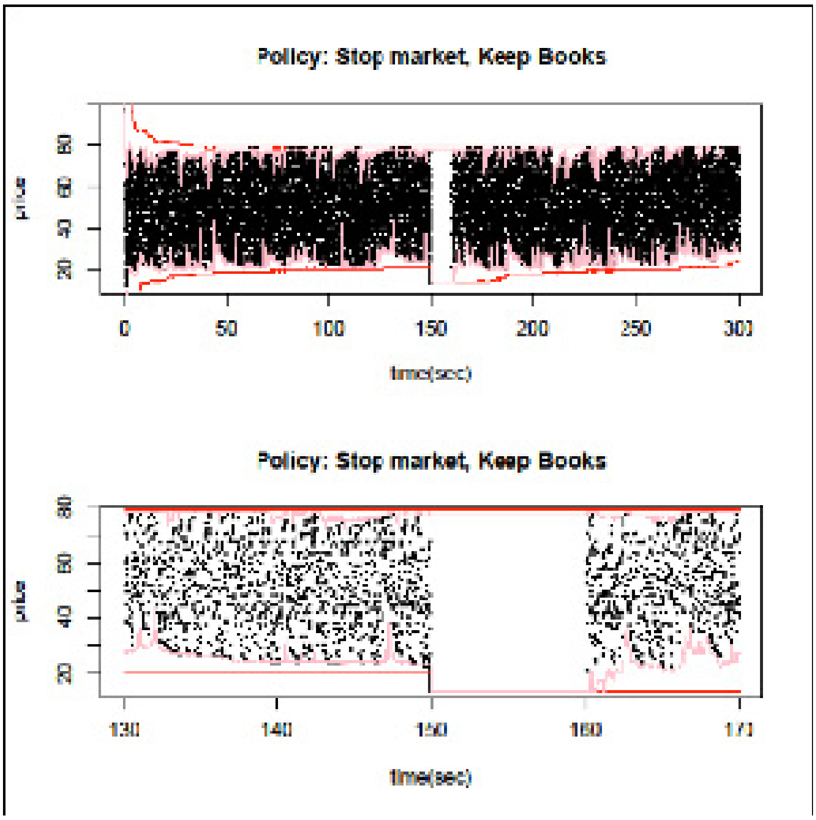
Figure 21. Waiting 10 sec with no orders or trades, then clearing books and resuming



4. Stop trading (10 seconds) and take no orders, keep books unchanged.

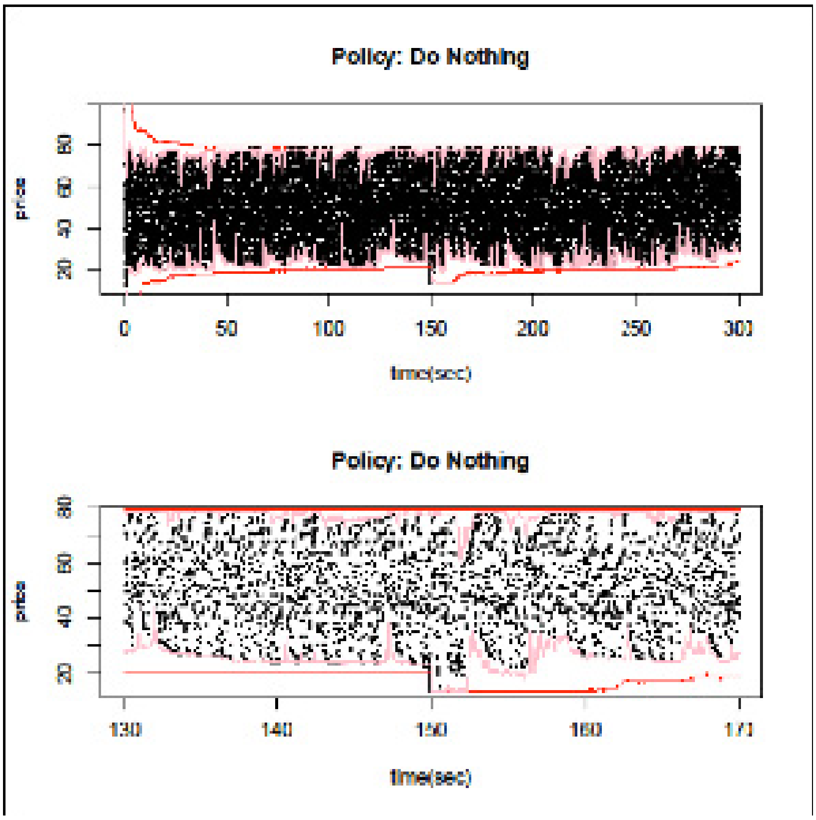
The policy here is to stop trading (10 seconds) as before, take no trades and facilitate no trade execution, keep all orders on the books and then resume trading. This maintains existing liquidity, but the asymmetry of liquidity created by the flash crash remains. As observed earlier it requires a minute or two for the balance of liquidity to be restored and the extra price variability caused by the hammer to be restored. See Figure 22.

Figure 22. Waiting 10 sec with no orders or trades, keeping books and resuming



5. We also provide Figure 23 for the market where nothing is done.

Figure 23. No treatment action performed



Remark

As elsewhere in this work, the effects on the book caused by the crash might be confounded by the behavior of individual traders which we do not model.

6. Future

We now provide our comments on what we expect to happen in the future.

6.1. Increased demand for market trading

6.2. Increased demand for rational, machine implemented regulations

- In the US, regulation NMS requires that an exchange fill orders in less than one second or forward the unfilled portion to another exchange

- To keep volume local, some exchanges are rewriting customer orders and creating flash orders, orders with rapid expiration, to lure mechanized traders into a deal.
[see http://static.seekingalpha.com/uploads/2009/7/26/saupload_flash_orders_diagram.jpg]

6.3. Technology

Currently:

- Modern cpus run about 10-50 gigaflops. That is 10 – 50 billion floating point operations per second, or 10-50 million floating point operations per ms (0.001 sec).

- Nvidia Corporation has pioneered graphics cards that are themselves high performance parallel computers, providing affordable graphics processing systems that can process 1000 gigaflops, or about 1 billion operations per millisecond. The effect of this is already being felt in games and scientific computing.

- A typical spinning hard drive still takes 5-10ms to locate a piece of data because the data is bound to a rotating platter of magnetic material that must physically rotate under a reader. In this time tens to hundreds of millions of basic calculations can take place. Memory based hard drives that do not involve a mechanical mechanism and have much lower access times are slowly replacing the spinning disk type, but are still relatively expensive.

- A data transfer between a local area network could take 0.1ms, within the EU, perhaps 30-80ms, from the US to the EU, 100ms or more. Similarly, in this time a large amount of calculations could take place.

- The world's largest publicly reported supercomputers are large-scale clusters of PC-type hardware running Linux. Large charitable efforts exist that use spare computing time of thousands of volunteers.

- There is a wide and growing array of inexpensive software available that can be used to create a program trading system and connect it to the exchanges.

- The Java Virtual Machine lets us run un-trusted programs in a web browser, and virtualization technology exists to run a Linux computer inside a Windows computer or vice versa.

- There are emerging markets for "on-demand" computing, such as Amazon's EC2.

- There is a common sense of an "information overload" in that people can search for or generate more data than they can analyze or comprehend.

In 10 years:

- Computers will continue to become faster and parallel computing more accessible, based solely on easier access to current developments.

- Secure containers, such as the Java Virtual Machine and Virtualization in general, will bring down the cost for exchanges of creating localized high frequency trading platforms for traders and other market participants that have direct access to market tick data at a higher frequency or bandwidth than will be physically possible for other market participants.

- Markets will exist where large amounts of unused computing time can be bought and sold on short notice, which could be called upon by traders as necessary. However, in the time required to communicate a full specification of market conditions to a remote analysis facility, it may well be that market conditions have changed. There will remain a competitive advantage for having compute facilities "close" to the exchanges.

- Memory based or solid state hard drives will be common, which will eliminate some of the bottleneck of retrieving and analyzing large amounts of data.

- It will still take time to retrieve data not in local storage, because the speed of light acts as the ultimate speed limit on internet communications. Because the speed of light is 300,000 km/sec, in 1ms an electrical signal could travel at most 300km. In practice, there are additional delays in the path.

- In the time it takes to retrieve a piece of data not in a local computer's memory, it will still be possible to do millions, perhaps billions, of elementary calculations.

- There will still be a sense of "information overload" from high frequency market systems in the sense that faster markets will produce more tick data than can be easily analyzed by regulators, or even fully transported around the network of market participants.

- These and other, unforeseen innovations will create regulatory issues, some of which involve bona-fide issues of fairness while others will involve incumbents wishing to protect an investment in an existing system or way of doing business.

- To integrate this hypothetical timeline with our current research, we note that liquidity is at its essence a list of buy and sell orders, which can exist either in memory, a hard drive, on a set of computers in a network, or in several computers across the internet. The trend towards faster local computing and faster local memory can alleviate flash crashes to the extent that more of the orders are available in the memory of a single computer responsible for matching orders in a market and can exacerbate flash crashes to the extent that order flow is removed from local memory to be stored elsewhere, or stored across a network or on multiple computers. The storage of orders elsewhere will create vulnerability inviting sophisticated algorithms to guess when a large order might disrupt the market before the liquidity can be recalled from the storage.

7. Summary

We use a simulation platform to study limit order markets with buy and sell orders with random values, arriving at random times. We induce a "flash crash" by means of a very large order, and analyze how it affects the liquidity of the book and the variability of the transaction prices. The simulations are performed under standard, continuous double-auction market microstructure, and under alternative structures, including varying order expiration times, 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.

7.1. Principles governing liquidity

Our view on liquidity in 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. Orders at various price points are themselves governed by a random process of uncoordinated decisions made by different bidders and based on different goals and different perceptions of the future.
4. 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; (ii) Erosion – asymmetric arrivals on one side of the book that have limits off the market, having the effect of reducing liquidity at several price points off the market.

7.2. Conclusions at a general level:

1. A flash flood can leave the market eroded.
2. Some market structures do a better job of fixing erosion than others.
3. Exact nature of crash and book erosion depends on the structure of the order flow.

7.3. Recommendation

While it may be helpful to require frequent traders to provide liquidity at all times, we do not recommend requiring minimal resting times. Rather, if there is a significant chance that a sudden fall in prices will have a long-term disruptive effect, we recommend switching to call auctions until the prices stabilize. However, we believe that most of the time the market needs no outside intervention to stabilize.

References

Alton, M. and Charles R. Plott. `` 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. September 2007.

Bossaerts, Peter, Charles R. Plott, William Zame, “Prices and Allocations in Financial Markets: Theory, Econometrics, and Experiments,” *Econometrica,* 75 (2007), 993-1038.

Coppejans, M., Domowitz, I, and Madhavan, A. (2004) Resiliency in an Automated Auction. Unpublished manuscript.

Cvitanic, J. and A. Kirilenko (2010) “High-Frequency Trading and Asset Prices”. Working paper.

Lee, B., Cheng, S-F., and Koh, A. (2010) “An Analysis of Extreme Price Shocks and Illiquidity Among Systematic Trend Followers.” *Review of Futures Markets* 18:4, pp.385-419.

Lee, B., Cheng, S-F., and Koh, A. (2011) Liquidity Withdrawal and the ``Flash Crash” on May 6, 2010. Working paper.

Madhavan, A. (1992) “Trading Mechanisms in Security Markets”. *Journal of Finance*, 47, pp. 607-641.

Plott, Charles R., and Vernon L. Smith (2005), “Handbook of Experimental Economics Results Vol. 1”, edited by C. Plott and V. L. Smith. Elsevier (2008).

Foresight
1 Victoria Street
London SW1H 0ET
www.foresight.gov.uk
URN: 12/1065