Algorithmic Finance 2 (2013) 213–231
DOI 10.3233/AF-13027
IOS Press

Dynamical trading mechanisms in limit order markets

Shilei Wang*
Department of Economics, Ca' Foscari University of Venice, 30121 Venice, Italy
**E-mail: shilei.wang@unive.it*

Abstract. This work's main purpose is to understand the price dynamics in a generic limit order market, and illustrate a dynamical trading mechanism that can be applied to explore its market microstructure. First and foremost, we capture the iterative nature of the limit order market, and quantitatively identify its capacities as a means to develop switching schemes for the appearances of different sorts of traders. After formally introducing a dynamical trading system to replace the complex limit order market, we then study trading processes in that trading system from both deterministic and stochastic perspectives, in the purpose of recognizing conditions of general instability and stochastic stability in the trading system. In the final part of this work, the dynamics of the spread and mid-price in a controlled trading system will be investigated, which fairly serves to verify the robustness of stochastic stability appearing in an uncontrolled trading system.

Keywords: Limit order market, dynamical trading system, controlled trading system, kernel region, buffering region, trading block, instability, stochastic stability

1. Introduction

This work aims to investigate and understand the price dynamics in a generic limit order market, in which the best quotes, as well as the induced spread and mid-price, are evidently important information carriers for both traders and analysts. Early studies of prices formation in these financial markets mainly concentrated on their economic natures, for instance, Demsetz (1968) considered the bid-ask spread as a markup "paid for predictable immediacy of exchange in organized markets" (Demsetz, 1968, p. 36), and regarded it as an important source in transaction costs. However, more recent studies started to explore the dynamics and evolutions appearing in these financial markets, partly because people realized that understanding the processes and dynamics seems very likely more important and more useful than merely explaining the equilibrium states. And such a tendency seems to be necessary as well, since the real financial markets created a few catastrophes in the past century, say in particular, the crash of October 1929, and the crash of October 1987 (*cf.* Sornette, 2003, pp. 5–7, 12–15). This work will also follow such a transformation, and, as we have put at the very beginning, study trading processes in the limit order market.

Most modern stock exchanges adopt electronic order-driven platforms, in which limit order books operate to match demand with supply, and shape value for time and liquidity. They surely provide opportunities to obtain stylized empirical observations of the limit order market. There are plenty of investigations published in past few years. For instance, Lehmann and Modest (1994) studied the trading mechanism and the liquidity in the Tokyo Stock Exchange, Biais, Hillion and Spatt (1995) studied the limit order book and the order flow in the Paris Bourse, Harris and Hasbrouck (1996) measured the performance of the SuperDOT traders in the NYSE, and finally Al-Suhaibani and Kryzanowski (2000) analyzed the order book and the order flow in the Saudi Stock Exchange.

As for theoretical understandings of the limit order market, the approaches of the existing studies might be roughly separated into two distinct categories, fortunately, which would also supply concrete cases in that methodological transformation as we mentioned before. Namely, one is to explain the performance of a limit order market and its stable states by modeling the behavior of traders in equilibrium. While the other is to consider the limit order market as a "super-trader" in order to either explain or predict its behavior statistically. The first category understands the strategic behavior of traders and generates testable implications by capturing some attributes of traders, for instance, being informed versus being uninformed as used by Kyle (1985), and also Glosten and Milgrom (1985), time preference used by Parlour (1998), and patience versus impatience used by Foucault, Kadana and Kandel (2005). The second category analyzes the limit order market by assuming there exist a few statistical laws for the dynamics or processes in the market. It is somehow able to catch certain profiles of the market, say notably, fat tails of the price distribution, concavity of the price impact function, and the scaling law of spread to orders (see Smith *et al.*, 2003, and Farmer, Patelli and Zovko, 2005).

This work will take an intermediary position between these two quite different methodologies. That's to say, we not only agree that the performance of a limit order market and its evolution can be determined by the behavior of rational traders involved in the market, but also keep in mind that statistical mechanics could be fairly important in its dynamics and process. Therefore, we suppose there is a large population of traders in a limit order market, and assume each trader in the population is rational, in the sense that her behavior is always strategically optimal in any game-theoretic framework. In practice, we assume that a rational trading decision in a very short time interval, as a quasi-static equilibrium, can be represented in terms of conformity with the optimal price distribution of the order book, which has a fine meaning of collective rationality. Evidently, the notions of individual rationality and collective rationality are thus supposed to be determined interchangeably.

In general, this work plans to study how individual traders affect the price dynamics in the limit order market, how different trading blocks influence the stability of the market, how a piece of randomness in a trading system can stochastically enhance its systemic stability, and why the market would evolve more predictably, in case we can control certain factors of the market. In brief, it aims to clarify a dynamical trading mechanism in the limit order market.

The writing of this work is organized into five sections. The first section is a general introduction to the literature, the methodology, the structure, and the awaiting questions. And the last section is mixed of a summary and some additional remarks on instability and stochastic stability.

In Section 2, we first construct the basic modeling framework, and introduce the atomic trading schemes as the necessary knowledge for traders in the limit order market. We then develop the switching laws for the appearances of different types of traders, and show that the market capacities of accepting limit-type and market-type traders can be measured by floor functions of the log-scaled spread, and similarly capacities of accepting buy-type and sell-type traders can be measured by floor functions of the log-scaled mid-price. These results are critical to the upcoming probabilistic setting of a random trading process.

In Section 3.1, we study deterministic trading processes in a dynamical trading system from a combinatorial perspective. We recognize sufficient conditions for its general instability, and identify the necessary condition for its stability — any trading process should contain at least one reducible trading block. In Section 3.2, we study stochastic trading processes with some certain probabilistic structures. We practically introduce two fundamental concepts — kernel region and buffering region — and show that the dynamical trading system will be stochastically stable in the kernel region, if its kernel region is moderately large and its buffering region is nonempty.

In Section 4, we check the robustness of stochastic stability for a regular uncontrolled trading system, by making its kernel region controlled to have some restricted properties. We show that the controlled dynamical trading system could be still stochastically stable, even if either its range of the spread or its range of the mid-price is extremely small. And thus, in a general sense, the stochastic stability of a regular trading system is robust.

2. Atomic Trading Scheme

2.1. Preliminary Framework

Consider a generic order-driven market with a large population of traders, whose attributes can be characterized by their trading directions and demands for the liquidity. As usual, the trading direction is either buying or selling initiation, while the liquidity demand determines the type of a submitted order, namely either a limit order or a market order.

The group of traders can thus be partitioned into four pairwise disjoint subgroups, each of which includes homogeneous traders, according to these two kinds of binary classification. It then appears that we have only four different types of traders:

(i) a buyer submitting a limit order,
(ii) a buyer submitting a market order,
(iii) a seller submitting a limit order,
(iv) a seller submitting a market order.

Let $\sigma(1)$, $\sigma(2)$, $\sigma(3)$, and $\sigma(4)$ denote these four types, respectively, where $\sigma : i \mapsto \sigma(i)$ is a normal permutation function defined on $\{1, 2, 3, 4\}$. Let the type space be Σ_4, then

$$\Sigma_4 = \{\sigma(1), \sigma(2), \sigma(3), \sigma(4)\}.$$

Evidently, any trader in the population must be of a unique type in Σ_4.

Assume that the depth of the limit order book is equal to the trading volume of any new order, thus a new transaction will either clear a limit order or add a more attractive limit order on the market, once there do not exist queued and hidden orders. So traders of any type in Σ_4 must affect the limit order book. We therefore call such traders *marginal traders*, in the sense that the (best) quotes will be definitely updated by their submitted orders.

Let the best bid, best ask, bid-ask spread, and mid-price in an order book be b, a, s, and m. The (best) bid-ask pair is denoted by (b, a) or $\mathbf{w}$. Note that $s = a - b$ and $m = (b + a)/2$. More generally, we define a *spread function*,

$$s : \mathbf{w} \mapsto s(\mathbf{w}),$$

such that $s(\mathbf{w}) = a - b$, and define a *mid-price function*,

$$m : \mathbf{w} \mapsto m(\mathbf{w}),$$

such that $m(\mathbf{w}) = (b + a)/2$, where $\mathbf{w} = (b, a)$.

Let the time domain be $\mathbb{Z}$. At each time $t \in \mathbb{Z}$, the time-dependent bid-ask pair will be denoted by (b_t, a_t) or $\mathbf{w}_t$. Once the time t is in the future, we introduce a two-dimensional random variable $\mathbf{D}_t$ to represent that unrealized bid-ask pair (B_t, A_t), where B_t and A_t are stochastic versions of b_t and a_t, respectively. Here, the normal notions of volatility and unpredictability could apply to B_t, A_t, and even $\mathbf{D}_t$.

Observe that there typically exists a tick size as the minimal change of prices in the market. Let $\tau > 0$ denote it, then we have $s > 0$, and moreover, $s \geq \tau$. In this work, we shall set a stricter condition for the lower bound of the bid-ask spread, that's to say, there is a lower bound $\underline{s} > \tau$, such that $s \geq \underline{s}$ at any time.

Naturally, there also exists an upper bound of the best ask a, in virtue of the limited value of any security for all trader. Let $\overline{a}$ denote it. Besides that, note that $b \geq 0$, otherwise, there would be no demand in the market, as the inverse of even the best bid would exist as a part of the ask side of the market. As a result, the bid-ask pair (b, a) should be located within a compact domain $W \subset \mathbb{R}^2$, which is defined by $b \geq 0$, $a \leq \overline{a}$, and $a - b \geq \underline{s}$. In the b-a plane, W can be geometrically represented by a triangle, whose vertices are $(0, \underline{s})$, $(0, \overline{a})$, and $(\overline{a} - \underline{s}, \overline{a})$, which has been shown in Fig. 1.

Suppose that the ordered quotes on both sides of the order book are naively distributed over $[0, b]$ and $[a, \overline{a}]$. In addition, we assume that the difference between adjacent quotes on the same side of the order book should be proportional to the bid-ask spread. This assumption as an empirical fact was found and statistically tested by Biais, Hillion and Spatt (1995) using data from the Paris Bourse, and also supported by Al-Suhaibani and Kryzanowski (2000) with evidence from the Saudi Stock Exchange. Furthermore, we set the difference of adjacent quotes equal to $\alpha(a - b)$ on the ask side of the book, and $\beta(a - b)$ on the bid side of the book, where $\alpha, \beta \in (0, 1)$.

The upper quote next to the best ask a is thus equal to $a^+ = a + \alpha(a - b)$, and the lower quote next to the best bid is equal to $b^- = b - \beta(a - b)$. Let's assume the ratio $\langle \beta, 1, \alpha \rangle$ of the differences between these four consecutive prices b^-, b, a, a^+ in the order book (see Fig. 2) is an indicator of optimal information aggregation or market efficiency in a very short time interval. That's to say, prices, which conform to such a ratio, suggest that the market should be in a quasi-static equilibrium, so that there is no profitable perturbation that could emerge in that short time interval. In practice,

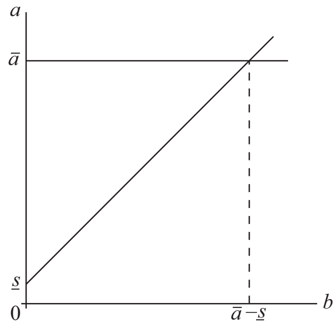

Fig. 1. The domain of the bid-ask pair.

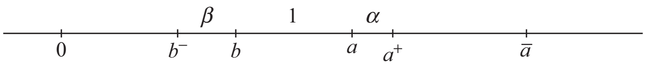

Fig. 2. Optimal structure of the limit order book.

the values of α and β can be roughly set as $1/2$, which was proposed by Biais, Hillion and Spatt (1995).

2.2. *Evolution of Bid-Ask Pair*

Suppose that a marginal trader gets to the market at the time t, just after the bid-ask pair (b_t, a_t) forms, thus she has all the information on the market, at and before the time t. Her trading decision at t can be denoted by a price, say p_t. If she submits a limit order, then $b_t < p_t < a_t$, while if she submits a market order, we have either $p_t \geq a_t$ or $p_t \leq b_t$. If she is a buyer, then $p_t \leq a_t$, while if she is a seller, then $p_t \geq b_t$. Recall that we have assumed the order book's depth at the quotes is equal to the trading volume of any new order. So a market order will definitely clear one of the two limit orders at the best quotes, and the cleared previous limit order will be replaced by a less attractive one. And a limit order will surely lower the bid-ask spread, and improve one of the best quotes.

Type $\sigma(1)$ If the marginal trader is a buyer and submits a limit order, then the best bid b_{t+1} at the time $t+1$ will be $p_t = b_t + \rho$, where $\rho > 0$, and the best ask a_{t+1} at the time $t+1$ will remain unchanged. Suppose the marginal trader's decision process can be described by her maximizing the utility function of p_t or equivalently ρ, say $u(p_t) = u(\rho + b_t)$, where u is concave in ρ. By the requirement of rationality, we have $\rho^* \in \operatorname{argmax}_\rho u(p_t)$, subject to $\rho^* > 0$ and $a_t - b_t - \rho^* \geq \underline{s}$. According to our assumption on the optimal structure of the limit order book in a very short time interval around t, ρ^* should comply with the ratio $\langle \beta, 1 \rangle$ on the bid side, which hence implies

$$\rho^* = \beta(a_t - b_t - \rho^*), \tag{1}$$

so $\rho^* = \frac{\beta}{1+\beta}(a_t - b_t)$. Thus $b_{t+1} = \frac{1}{1+\beta} b_t + \frac{\beta}{1+\beta} a_t$, and $a_{t+1} = a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_1$, where

$$S_1 = \begin{pmatrix} \frac{1}{1+\beta} & 0 \\ \frac{\beta}{1+\beta} & 1 \end{pmatrix},$$

and $\mathbf{w}_{t+1}$ is the realization of the random variable $\mathbf{D}_{t+1}$.

After this trade of limit order on the bid side, the spread $s_t = a_t - b_t$ and mid-price $m_t = (b_t + a_t)/2$ will be updated respectively to

$$s_{t+1} = \frac{1}{1+\beta} s_t,$$

$$m_{t+1} = m_t + \frac{\beta}{2(1+\beta)} s_t.$$

Note that $s_{t+1} \geq \underline{s}$, so $s_t \geq (1+\beta)\underline{s}$. The marginal trader is of type $\sigma(1)$, only if $s_t \geq (1+\beta)\underline{s}$. And if she is of type $\sigma(1)$, she will choose an optimal improvement $\rho^* \geq \beta\underline{s}$.

Type $\sigma(2)$ If the marginal trader is again a buyer, but now she submits a market order hitting a_t, then the best ask a_{t+1} at the next period will be $a_t + \alpha(a_t - b_t)$, and the best bid b_{t+1} will remain same as b_t. Here, her decision process is overlaid with the principle of price-time priority employed in the limit order market. So her decision set is the singleton $\{a_t\}$, which then implies that $p_t = a_t$. Consequently, we have $b_{t+1} = b_t$ and $a_{t+1} = -\alpha b_t + (1+\alpha) a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_2$, where

$$S_2 = \begin{pmatrix} 1 & -\alpha \\ 0 & 1+\alpha \end{pmatrix}.$$

After this trade of market order on the bid side, the new spread and mid-price will be

$$s_{t+1} = (1+\alpha) s_t,$$

$$m_{t+1} = m_t + \frac{\alpha}{2} s_t.$$

Note that $s_{t+1} \leq \overline{a}$, so $s_t \leq \overline{a}/(1+\alpha)$. The marginal trader is of type $\sigma(2)$, only if $s_t \leq \overline{a}/(1+\alpha)$.

Type $\sigma(3)$ If the marginal trader is a seller, and she submits a limit order, then the best ask a_{t+1} at the time $t+1$ will be $a_t - \theta$, where $\theta > 0$, and the best bid b_{t+1} at the time $t+1$ will remain same as b_t. Similar to the type $\sigma(1)$, this type of marginal trader's decision process can be described as maximizing her utility $v(p_t) = v(-\theta + a_t)$, where v is convex in θ. Her decision will admit the optimal choice $\theta^* \in \operatorname{argmax}_\theta v(p_t)$, subject to $\theta^* > 0$ and $a_t - \theta^* - b_t \geq \underline{s}$.

By the assumption that the optimal ratio $\langle 1, \alpha \rangle$ on the ask side represents the quasi-static equilibrium around the time t in the market, we have

$$\theta^* = \alpha(a_t - \theta^* - b_t), \tag{2}$$

so $\theta^* = \frac{\alpha}{1+\alpha}(a_t - b_t)$. Thus $b_{t+1} = b_t$, and $a_{t+1} = \frac{\alpha}{1+\alpha} b_t + \frac{1}{1+\alpha} a_t$. Or more concisely, $\mathbf{w}_{t+1} = \mathbf{w}_t S_3$, where

$$S_3 = \begin{pmatrix} 1 & \frac{\alpha}{1+\alpha} \\ 0 & \frac{1}{1+\alpha} \end{pmatrix}.$$

After this trade of limit order on the ask side, the spread and mid-price will be updated to

$$s_{t+1} = \frac{1}{1+\alpha} s_t,$$
$$m_{t+1} = m_t - \frac{\alpha}{2(1+\alpha)} s_t.$$

Similar to the type $\sigma(1)$, we have $s_t \geq (1+\alpha)\underline{s}$, since s_{t+1} can not be less than $\underline{s}$. The marginal trader is of type $\sigma(3)$, only if $s_t \geq (1+\alpha)\underline{s}$. And if she is of type $\sigma(3)$, she will choose an optimal improvement $\theta^* \geq \alpha\underline{s}$.

Type $\sigma(4)$ If the marginal trader is a seller and submits a market order hitting b_t, then the best bid b_{t+1} at the time $t+1$ will be $b_t - \beta(a_t - b_t)$, and the best ask a_{t+1} will remain unchanged as a_t. Her decision is restricted to choosing p_t to maximize her utility subject to $p_t \in \{b_t\}$, so the optimal choice is $p_t = b_t$. We have $b_{t+1} = (1+\beta)b_t - \beta a_t$ and $a_{t+1} = a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_4$, where

$$S_4 = \begin{pmatrix} 1+\beta & 0 \\ -\beta & 1 \end{pmatrix}.$$

After this trade of market order on the ask side, we will have

$$s_{t+1} = (1+\beta)s_t,$$
$$m_{t+1} = m_t - \frac{\beta}{2} s_t.$$

Similar to the type $\sigma(2)$, we should have $s_t \leq \overline{a}/(1+\beta)$. And only if $s_t \leq \overline{a}/(1+\beta)$, the marginal trader could be of type $\sigma(4)$.

Type Space Observe that any spread s can increase to $(1+\alpha)s$ or $(1+\beta)s$, and decrease to $s/(1+\alpha)$ or $s/(1+\beta)$ after a new order, so the maximal difference generated by α and β could be $|\alpha - \beta|s$. Since $\overline{a} \gg \underline{s}$ in most normal limit order markets, and both α and β are roughly close to $1/2$, $|\alpha - \beta|s$ will be extremely small in regard to $\overline{a} - \underline{s}$. Therefore, we can, without loss of generality, assume $\alpha = \beta$ to set $|\alpha - \beta|s$ exactly equal to 0. The ratio of $\langle \beta, 1, \alpha \rangle$ will then be replaced by the simpler one, $\langle \alpha, 1, \alpha \rangle$.

Consider an arbitrary initial bid-ask pair (b, a). If a marginal trader of type $\sigma(1)$ comes to the market, the bid-ask pair in the next period will be (b^+, a), where $b^+ > b$. If the marginal trader is of type $\sigma(2)$, it will be (b, a^+), where $a^+ > a$. If the marginal trader is of type $\sigma(3)$, it will be (b, a^-), where $a^- < a$. Finally, if the marginal trader is of type $\sigma(4)$, it will be (b^-, a), where $b^- < b$. Note that

$$b^+ - b = a - a^- = \frac{\alpha}{1+\alpha} s,$$
$$a - b^+ = a^- - b = \frac{s}{1+\alpha},$$

and also

$$a^+ - a = b - b^- = \alpha s,$$
$$a^+ - b = a - b^- = (1+\alpha)s.$$

The types $\sigma(2)$ and $\sigma(4)$ will cause a greater bid-ask spread than s, namely $(1+\alpha)s$, while the types $\sigma(1)$ and $\sigma(3)$ will cause a smaller bid-ask spread than s, namely $s/(1+\alpha)$. In fact, $\sigma(2)$ and $\sigma(4)$ are both market-type, while $\sigma(1)$ and $\sigma(3)$ are both limit-type (see Fig. 3).

By similar computations as above, we can obtain

$$a + b^+ = 2m + \frac{\alpha}{1+\alpha} s,$$
$$a^+ + b = 2m + \alpha s,$$

and also

$$a^- + b = 2m - \frac{\alpha}{1+\alpha} s,$$
$$a + b^- = 2m - \alpha s.$$

The types $\sigma(1)$ and $\sigma(2)$ will generate a new mid-price greater than m, while the types $\sigma(3)$ and $\sigma(4)$ will generate a new mid-price smaller than m. We clearly see that $\sigma(1)$ and $\sigma(2)$ are both buy-type, while $\sigma(3)$ and $\sigma(4)$ are both sell-type (see again Fig. 3).

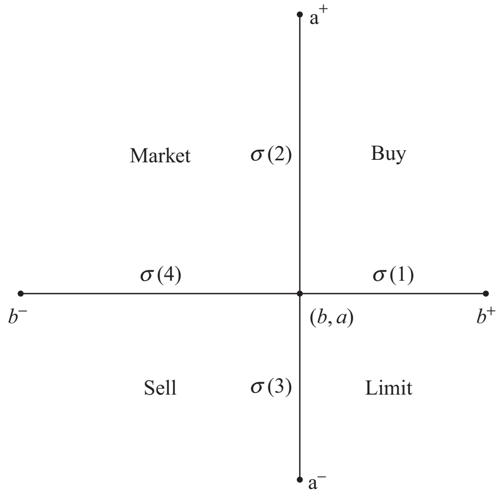

Fig. 3. Partition of the type space.

2.3. *Capacity and Switching Law*

From the discussions on trading decisions of different marginal traders, we can see how the bid-ask spread and mid-price develop in an iterative way. Notice that a limit order will change the bid-ask spread from s to $s/(1+\alpha)$, while a market order will change it from s to $(1+\alpha)s$. It appears that any s will converge to infinity after sufficiently many market orders, and any s will converge to zero after infinitely many limit orders. Recall that s should be bounded within the interval $[\underline{s}, \overline{a}]$ (a line $a = b + s$ should pass W, see Fig. 1), so when s is too close to its bounds, the market should have stability incentives to make s bounce away against them. The mid-price m has a similar internal stability scheme, as m should be bounded within the interval $[\underline{s}/2, \overline{a} - \underline{s}/2]$ (a line $a + b = 2m$ should pass W, see also Fig. 1). Quite intuitively, we shall say such internal stability schemes originate in some hidden "gravitational forces" in the market, in the sense that they can control the appearances of different types of orders, and determine their switching possibilities.

We should admit that the force related to the bid-ask spread and that related to the mid-price should have different actions on the market, but similar analytical natures. In consideration of such facts, we will mainly study the force related to the bid-ask spread, and yet we will directly state similar results on the force related to the mid-price at the end of this part.

To investigate the gravitational force for the bid-ask spread, we first concentrate on such spreads that are very close to either $\underline{s}$ or $\overline{a}$, so that we could easily see its working principles. If the spread is sufficiently great, the hidden force will attract limit-type traders, $\sigma(1)$ and $\sigma(3)$, and repel market-type traders, $\sigma(2)$ and $\sigma(4)$, but of course doesn't necessarily reject them until the spread is close enough to the upper bound $\overline{a}$. On the other hand, if the bid-ask spread is extremely small, then such a hidden force will attract market-type traders and repel limit-type traders, and it will reject limit-type traders once the spread is close enough to $\underline{s}$.

Formally stating, the market with a spread s only accepts traders of type $\sigma(2)$ and type $\sigma(4)$, if

$$\underline{s} \le s < (1+\alpha)\underline{s}.$$

Suppose there exists a nonempty interval $((1-\gamma)\overline{a}, \overline{a}]$, where $0 < \gamma < 1$ and $(1-\gamma)(1+\alpha) \le 1$[1], or equivalently,

$$\alpha/(1+\alpha) \le \gamma < 1,$$

under which the market will never accept any more market-type trader. Thus the market with a spread s only accepts traders of type $\sigma(1)$ and type $\sigma(3)$, if

$$(1-\gamma)\overline{a} < s \le \overline{a}.$$

In the remaining interval,

$$(1+\alpha)\underline{s} \le s \le (1-\gamma)\overline{a},$$

the market can accept traders of any type in Σ_4. In fact, $[\underline{s}, \overline{a}]$ is now partitioned into three pairwise disjoint intervals, as shown in Fig. 4.

We then establish a result stating that the capacity of accepting limit-type traders in the limit order market can be measured by a function of the bid-ask spread.

Proposition 2.1. The maximal number of limit-type traders, who can be continuously accepted by a limit order market with a spread s, is determined by the function,

$$z(s) = \left\lfloor \frac{\log s - \log \underline{s}}{\log(1+\alpha)} \right\rfloor. \tag{3}$$

Proof. Define a sequence of consecutive intervals,

$$[(1+\alpha)^i \underline{s}, (1+\alpha)^{i+1} \underline{s}),$$

[1]This inequality gives a necessary condition for a zero-capacity of accepting market-type traders. If we assume $(1-\gamma)(1+\alpha) = 1$, although we could have less parameters, there would be also less room for interesting analysis.

Fig. 4. Acceptable traders in the market.

for $i \in \{0, 1, \ldots, n\}$, and

$$n = \max\{i \in \mathbb{Z} : (1+\alpha)^i \leq \overline{a}\} - 1.$$

For all $s \in [\underline{s}, (1+\alpha)^{n+1}\underline{s})$, there is a unique $j(s) \in \{0, 1, \ldots, n\}$, such that

$$s \in [(1+\alpha)^{j(s)}\underline{s}, (1+\alpha)^{j(s)+1}\underline{s}).$$

We want to show that $z(s) = j(s)$ by induction. If $j(s) = 0$, then $s \in [\underline{s}, (1+\alpha)\underline{s})$, and the limit order market will reject limit-type orders, so $z(s) = 0$. Assume $z(s) = j(s)$ is true for all $j(s) \leq k$, and consider $j(s) = k+1$ such that $s \in [(1+\alpha)^{k+1}\underline{s}, (1+\alpha)^{k+2}\underline{s})$. After a marginal trader of type $\sigma(1)$ or type $\sigma(3)$ comes to the market, s will be updated to $s' = s/(1+\alpha) \in [(1+\alpha)^k\underline{s}, (1+\alpha)^{k+1}\underline{s})$. By the assumption, we know $z(s') = k$, so $z(s) = z(s') + 1 = k+1 = j(s)$.

If $s \in [(1+\alpha)^{n+1}\underline{s}, \overline{a}]$, then $s \in [(1+\alpha)^{n+1}\underline{s}, (1+\alpha)^{n+2}\underline{s})$, as $(1+\alpha)^{n+2}\underline{s} > \overline{a}$. Here $j(s) = n+1$, so we have $z(s) = n+1 = j(s)$ by induction, simply as $z(s') = j(s')$ for all $j(s') = n$.

Therefore, $z(s)$ satisfies

$$(1+\alpha)^{z(s)}\underline{s} \leq s < (1+\alpha)^{z(s)+1}\underline{s},$$

and thus we have

$$z(s) \leq \frac{\log s - \log \underline{s}}{\log(1+\alpha)} < z(s) + 1,$$

which exactly defines the floor function. □

We can thus state that there is an exponential law of the bid-ask spread s to the market's capacity of accepting limit-type traders n,

$$s = (1+\alpha)^n \underline{s}, \tag{4}$$

where $n \in [z(s), z(s)+1]$.

To a certain extent, the appearance probability of a new limit-type trader should be simply determined by its capacity of accepting limit-type traders. Since the capacity is a function of the spread, that probability should be also a function of the spread. Let $f : [\underline{s}, \overline{a}] \to [0, 1]$ be such a function[2]. Since the log-scaled spread has been used to measure the capacity, $f(s)$ is actually a function of $\log s$, and moreover, we suppose $f(s)$ is positively linear in $\log s$. Thus we could define $f(s)$ more precisely in the following way. If $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, then

$$f(s) = k_1 \log s + k_2, \tag{5}$$

where $k_1 \geq 0$ and k_2 are constants depending on $\underline{s}, \overline{a}, \alpha, \gamma$. If $s \in [\underline{s}, (1+\alpha)\underline{s})$, then $f(s) = 0$. And if $s \in ((1-\gamma)\overline{a}, \overline{a}]$, then $f(s) = 1$.

Similar to Proposition 2.1, a result on the capacity of accepting market-type traders in the market can be proposed, and yet its proof will not be provided, as its logic is nearly the same as that of Proposition 2.1.

Proposition 2.2. The maximal number of market-type traders, who can be continuously accepted by a limit order market with a spread s, is determined by the function,

$$y(s) = \left\lfloor \frac{\log((1-\gamma)\overline{a}) - \log s}{\log(1+\alpha)} + 1 \right\rfloor^+, \tag{6}$$

where $\lfloor x \rfloor^+ = \max\{\lfloor x \rfloor, 0\}$.

Evidently, there is also an exponential law of the bid-ask spread s to the market's capacity of accepting market-type traders n,

$$s \propto (1+\alpha)^{-n}\overline{a}, \tag{7}$$

where $n \in [y(s), y(s)+1]$.

The appearance probability function $g : [\underline{s}, \overline{a}] \to [0, 1]$, which specifies the probability of a new market-type trader appearing in the market with a spread s, can be defined as a negatively linear function of $\log s$. Concretely, if $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, then

$$g(s) = k_3 \log s + k_4, \tag{8}$$

[2]Note that $f(s)$ is not a probability measure over the spread domain $[\underline{s}, \overline{a}]$, as we are not considering the uncertainty in the spread, but the uncertainty in the appearance of different types of traders at any deterministic spread s.

where $k_3 \leq 0$ and k_4 are again constants determined by $\underline{s}, \overline{a}, \alpha, \gamma$. If $s \in ((1-\gamma)\overline{a}, \overline{a}]$, then $g(s) = 0$. And if $s \in [\underline{s}, (1+\alpha)\underline{s})$, then $g(s) = 1$.

Note that, in any case, a new trader appearing in the market is either limit-type or market-type, that's to say, we have $f(s) + g(s) = 1$ for all s. Intuitively, the probabilities $f(s)$ as well as $g(s)$ at any s can be thought of to be a measure of switching possibility between the limit-type and market-type traders in the market. We can thus rigorously state such a switching law between limit type and market type as follows:

(i) If $s \in [\underline{s}, (1+\alpha)\underline{s})$, the probability of switching from limit type to market type is 1, and that of switching from market type to limit type is 0.
(ii) If $s \in ((1-\gamma)\overline{a}, \overline{a}]$, the probability of switching from market type to limit type is 1, and that of switching from limit type to market type is 0.
(iii) If $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, the probability of switching from limit type to market type is decreasing according to $g(s)$, while the probability of switching from market type to limit type is increasing according to $f(s)$.

With regard to the mid-price as the counterpart of the spread, we can also develop, through a highly similar reasoning process, the switching law between buy type (*i.e.*, type $\sigma(1)$ and $\sigma(2)$) and sell type (*i.e.*, type $\sigma(3)$ and $\sigma(4)$). Once again, we partition the mid-price domain $[\underline{s}/2, \overline{a}-\underline{s}/2]$ into three pairwise disjoint intervals,

$$[\underline{s}/2, (1+\delta)\underline{s}/2),$$

$$[(1+\delta)\underline{s}/2, (1-\epsilon)(\overline{a}-\underline{s}/2)],$$

$$((1-\epsilon)(\overline{a}-\underline{s}/2), \overline{a}-\underline{s}/2],$$

where $\delta > 0$, and $\delta/(1+\delta) \leq \epsilon < 1$, as we need initially assume $0 < \epsilon < 1$ and $(1+\delta)(1-\epsilon) \leq 1$. In short, we shall directly state that switching law between buy type and sell type as follows:

(i) If $m \in [\underline{s}/2, (1+\delta)\underline{s}/2)$, sell type will switch to buy type for sure.
(ii) If $m \in ((1-\epsilon)(\overline{a}-\underline{s}/2), \overline{a}-\underline{s}/2]$, buy type will switch to sell type for sure.
(iii) In the remaining domain of m, the probability of switching from buy type to sell type is increasing with $\log m$, and that of switching from sell type to buy type is decreasing with $\log m$.

3. Iterated Trading Process

3.1. Sequential Trading

Define four linear functions mapping W to itself,

$$f_i(\mathbf{w}) = \mathbf{w}S_i,\ i \in \{1,2,3,4\},$$

where S_1, S_2, S_3, S_4 are 2×2 matrices as defined in Section 2.2. Recall that β in S_1 and S_4 are now replaced by α. Let F be the collection of these four functions, then $F = \{f_1, f_2, f_3, f_4\}$.

For each $i \in \{1,2,3,4\}$, and any given $\mathbf{w} \in W$, define a convex set,

$$L_i(\mathbf{w}) = \{\lambda\mathbf{w} + (1-\lambda)\mathbf{w}S_i : 0 \leq \lambda \leq 1\}.$$

Notice that $L_i(\mathbf{w})$ is actually the line segment between $\mathbf{w}$ and $\mathbf{w}S_i$ in the b-a plane. More precisely, we redefine $f_i(\mathbf{w})$ to be

$$f_i(\mathbf{w}) \in \begin{cases} \{\mathbf{w}S_i\} & \text{, if } \mathbf{w}S_i \in W \\ L_i(\mathbf{w}) \cap \partial W, & \text{if } \mathbf{w}S_i \notin W \end{cases}$$

where ∂W denotes the boundary of the closed domain W. Since both $\{\mathbf{w}S_i\}$ and $L_i(\mathbf{w}) \cap \partial W$ are singletons, $f_i(\mathbf{w})$ will take either the value of $\mathbf{w}S_i$ or the unique element in $L_i(\mathbf{w}) \cap \partial W$, hence it is essentially a well-defined function.

We should notice that the domain W will then have an absorbing barrier, in such a sense that the dynamics induced by any $f_i \in F$ will be always bounded within W. To state the sharp distinction between a state absorbed on ∂W and a state in $W \setminus \partial W$, we shall propose two useful notions to describe the states in ∂W. If the state of a limit order market is first absorbed on ∂W at a time t, we say the market is *unstable* before the time t, and it is in a *crash* or *catastrophe* at and after the time t.

Definition 3.1. The pair (W, f_i) is called a *trading system* generated by the trader of type $\sigma(i)$ for all $i \in \{1,2,3,4\}$.

Each trading system (W, f_i) for $i \in \{1,2,3,4\}$ can create a certain dynamics of bid-ask pairs in the domain W. For any initial state $\mathbf{w} \in W$, the bid-ask pair in the trading system (W, f_i) will eventually hit the point $\mathbf{w}_i \in \partial W$, where $\mathbf{w}_1 = (a - \underline{s}, a)$, $\mathbf{w}_2 = (b, \overline{a})$, $\mathbf{w}_3 = (b, b+\underline{s})$, and $\mathbf{w}_4 = (0, a)$. These generic states are also shown geometrically in the b-a plane (see Fig. 5).

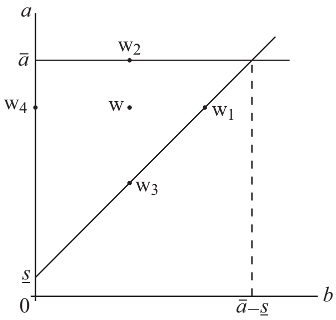

Fig. 5. The hitting states absorbed on ∂W.

Proposition 3.2. If a fixed-type marginal trader repeatedly comes to a limit order market, then the market starting from any initial state in W is unstable.

Proof. Consider any initial state $\mathbf{w} = (b, a)$ in W. After n forward periods with the marginal trader of type $\sigma(1)$, the bid-ask pair will be $\mathbf{w}S_1^n$, which will converge to (a, a), if n goes to infinity. But in the trading system (W, f_1), the bid-ask pair should always stay in W, thus the last bid-ask pair remaining in the trading system is $(a - \underline{s}, a) \in \partial W$. It means that the trading system will monotonically move to a crash, and hence it is unstable. Similar arguments can be made for the other three types, and they will complete the proof. □

In Proposition 3.2, we actually consider a countably infinite sequence of marginal traders with a constant type, $\{q, q, \ldots\}$, for all $q \in \Sigma_4$. It appears to us that the trading system involved with $\{q, q, \ldots\}$ must be unstable. In general, we can consider a countably infinite sequence of marginal traders with variable types, $\{q_t,\, t \in \mathbb{Z}\}$, where $q_t \in \Sigma_4$ for all $t \in \mathbb{Z}$, and investigate the stability of the trading system involved with $\{q_t,\, t \in \mathbb{Z}\}$. Evidently, the set of all such $\{q_t,\, t \in \mathbb{Z}\}$ can be denoted by Σ_4^ω.

If the type of marginal trader at a time t is $q_t = \sigma(i)$, the atomic trading scheme functioning at that time will be f_i for all $i \in \{1, 2, 3, 4\}$. Thus the permutation function σ can relate the functioning scheme $f_i \in F$ to the marginal trader of type $\sigma(i) \in \Sigma_4$.

Definition 3.3. The triplet (W, F, σ) is called a *dynamical trading system* or an *iterated trading system*.

Note that (W, f_i) is a discrete dynamical system, and (W, F) is essentially an iterated function system. In our definition of dynamical trading system, σ is introduced additionally to determine the functioning schemes in (W, F) for all sequence of marginal traders in Σ_4^ω. Recall that F has a parameter $\alpha \in (0, 1)$, so the dynamical trading system (W, F, σ) depends on α as well.

First of all, we are interested in identifying certain trading blocks in a sequence of traders that will never change the state of the dynamical trading system (W, F, σ). Notice that $S_1 S_4 = I$ and $S_2 S_3 = I$ for all $\alpha \in (0, 1)$, where I denotes the identity matrix of order 2, thus $\{\sigma(1), \sigma(4)\}$, $\{\sigma(4), \sigma(1)\}$, $\{\sigma(2), \sigma(3)\}$, and $\{\sigma(3), \sigma(2)\}$ are all such trading blocks for all $\alpha \in (0, 1)$.

Definition 3.4. A *periodic block* is a consecutive trading block that will not change any bid-ask pair in a specific dynamical trading system.

Notice that a combination of periodic blocks is again a periodic block. For instance,

$$\{\sigma(1), \sigma(4), \sigma(2), \sigma(3)\}$$

is a periodic block for all $\alpha \in (0, 1)$, as $\{\sigma(1), \sigma(4)\}$ and $\{\sigma(2), \sigma(3)\}$ are both general periodic blocks. So we need to catch the invariant part of periodic block.

Definition 3.5. A periodic block is called *minimal*, if it has no proper subtuple that is again a periodic block.

Any periodic block can be reduced into a series of minimal ones. Note that a periodic block C is either minimal or not. If C is minimal, it is equivalent to itself. If C is not minimal, we can always find a proper subtuple $C' \subset C$ such that both C' and $C \setminus C'$ are still periodic blocks. We can eventually have a series of minimal periodic blocks by applying this process recursively.

Example 3.6. If $\alpha \in (0, 1)$, the periodic block

$$\{\sigma(1), \sigma(2), \sigma(3), \sigma(4)\}$$

has two minimal ones, $\{\sigma(2), \sigma(3)\}$ and $\{\sigma(1), \sigma(4)\}$, while the minimal periodic blocks of

$$\{\sigma(4), \sigma(1), \sigma(1), \sigma(4), \sigma(4), \sigma(1)\}$$

are $\{\sigma(1), \sigma(4)\}$, and double $\{\sigma(4), \sigma(1)\}$.

If $\alpha = 1/2$, the consecutive trading block

$$\{\sigma(2), \sigma(1), \sigma(2), \sigma(1), \\ \sigma(3), \sigma(4), \sigma(3), \sigma(4), \sigma(3), \sigma(4)\}$$

is periodic, and it is also minimal.

If $\alpha = 1/3$,

$$\{\sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(4), \sigma(3), \sigma(4), \sigma(3), \sigma(4), \sigma(3)\}$$

is a minimal periodic block.

Lemma 3.7. The number of marginal traders in any minimal periodic block is finite and even.

Proof. Consider a minimal periodic block C, and assume the number of traders in C is infinite. Then C must pass infinite bid-ask pairs. If not, we suppose C passes finite bid-ask pairs. Since the number of traders in C is infinite, there must exist a closed route, such that a related subset of C is a periodic block, which contradicts that C is minimal.

Assume the initial condition of C is $\mathbf{w} \in W$ with a bid-ask spread s. Note that s can be updated into either $(1+\alpha)s$ or $s/(1+\alpha)$, so any bid-ask pair on the trajectory will have a spread in the set

$$S_{\mathbf{w}} = \{(1+\alpha)^i s : -N_1 \leq i \leq N_2 \text{ and } i \in \mathbb{Z}\},$$

where $N_1, N_2 \in \mathbb{Z}^+$ are finite, as W is bounded. Let the bid-ask pair $\mathbf{w}_r$ be the first state with a spread r on the trajectory starting from $\mathbf{w}$, for all $r \in S_{\mathbf{w}}$, where $\mathbf{w}_r = \mathbf{w}$, if $r = s$. $\mathbf{w}_r$ will be updated to $\mathbf{w}_r \pm (\alpha r/(1+\alpha), \alpha r/(1+\alpha))$ by the block $\{\sigma(1), \sigma(2)\}$ or $\{\sigma(3), \sigma(4)\}$, and to $\mathbf{w}_r \pm (\alpha r, \alpha r)$ by the block $\{\sigma(2), \sigma(1)\}$ or $\{\sigma(4), \sigma(3)\}$. So at the constant-spread line $a - b = r$, all the possible states on the trajectory have the form,

$$\mathbf{w}_r + k_1(\alpha r, \alpha r) + k_2\Big(\frac{\alpha}{1+\alpha}r, \frac{\alpha}{1+\alpha}r\Big) = \mathbf{w}_r + \Big(k_1\alpha + \frac{k_2\alpha}{1+\alpha}, k_1\alpha + \frac{k_2\alpha}{1+\alpha}\Big)r,$$

where $k_1, k_2 \in \mathbb{Z}$ and they are finite, as W is bounded. Since α, k_1, k_2 are finite, all the possible states with a given spread r on the trajectory are finite, and hence all the states in W starting from $\mathbf{w}$ are finite. So C can not pass infinite bid-ask pairs, which implies the number of marginal traders in C must be finite.

Suppose C has $2n+1$ traders, where $n \in \mathbb{Z}^+$, and assume it will pass x different bid-ask pairs, the collection of which is denoted by the set P. By Proposition 3.2, any trader of type $\sigma(i)$ will definitely update the bid-ask pair in a trading system (W, f_i), so $1 < x < \infty$. Let P be the set of nodes in a graph, so any trader in C will link two different nodes in P. Since there are $2n+1$ traders, we have $2n+1$ links in this graph. But if there exists a directed circle, such that the bid-ask pair after this block will not be changed, then the number of links of any node in P should be even. So the total links in this graph should be even, which contradicts that the number of links is $2n+1$. Therefore, a block with $2n+1$ traders can not be periodic, which completes the proof. □

Note that we can have an equivalent reduced sequence of traders by identifying and then removing (minimal) periodic blocks iteratively in any sequence of traders, as we just delete some closed routes of bid-ask pairs, which will not change the dynamics in the dynamical trading system as a whole.

Definition 3.8. A sequence of marginal traders is called *irreducible*, if it contains no minimal periodic block.

Proposition 3.9. A limit order market that accepts any irreducible sequence of traders is unstable.

Proof. Note that any market functioning for infinite periods must contain several minimal periodic blocks, otherwise it is a minimal periodic block with infinite traders, which contradicts Lemma 3.7. Since any irreducible sequence of marginal traders contains no minimal periodic block, the number of marginal traders in any irreducible sequence must be finite, otherwise we have a market functioning for infinite periods has no minimal periodic block. If the number of marginal traders in a sequence is finite, then the market must function only for finite periods. So the bid-ask pair in the market must be absorbed on ∂W, and hence the market will be in a crash, which implies that the market must be unstable. □

It is clear that any market functioning for infinite periods will never accept an irreducible sequence of marginal traders. Or we can say the sequence of marginal traders in a stable market should be infinite and reducible, so that we can always find some minimal periodic blocks that stay in the market for finite periods.

A similar concept to periodic block is the well-known notion of hedge, as the role of risk sharing through assets diversification has a counterpart here, namely, instability sharing through orders grouping in the limit order market. Our result suggests that periodic blocks as the "hedging" units in a limit order market should be necessary for its dynamic stability.

3.2. Stochastic Trading

In the last subsection, we studied a dynamical trading system from a combinatorial perspective. In fact, we considered all the possible permutations of countably infinite marginal traders, which form the space Σ_4^ω. We find two general categories of sequences of marginal traders in Σ_4^ω, which can sufficiently cause an unstable limit order market. That's to say, the sequence of marginal traders with a constant type, as stated in Proposition 3.2, and any irreducible sequence that contains no minimal periodic block, as stated in Proposition 3.9. Moreover, if we realize that a sequence of marginal traders with a constant type is certainly irreducible, Proposition 3.2 would then become a natural corollary of Proposition 3.9 at this stage.

In this part, we will take a different perspective to study sequential trading processes in the limit order market. We assume there is a certain probability structure over the space Σ_4^ω, and hence the dynamics of bid-ask pairs in the dynamical trading system (W, F, σ) will become random. Not surprisingly, the related limit order market can be stochastically stable, in the sense that the random trajectory in (W, F, σ) will not be absorbed on ∂W almost surely, in other words, the market will not be in a crash almost surely.

Assume the appearance probability of any type in Σ_4 at any state $\mathbf{w} \in W$ is stationary, *i.e.*, independent of the time. Define a function

$$\pi : W \to [0,1]^4,$$

such that, at any state $\mathbf{w} \in W$, $\pi(\mathbf{w})$ is the 4-tuple of the appearance probabilities of type $\sigma(1)$, type $\sigma(2)$, type $\sigma(3)$, and type $\sigma(4)$, or

$$\pi(\mathbf{w}) = \big(\pi_1(\mathbf{w}), \pi_2(\mathbf{w}), \pi_3(\mathbf{w}), \pi_4(\mathbf{w})\big),$$

with $\sum_{i=1}^4 \pi_i(\mathbf{w}) = 1$, where $\pi_i(\mathbf{w})$ denotes the appearance probability of type $\sigma(i)$ at the state $\mathbf{w}$.

We can therefore have four aggregated appearance probability functions, which can be directly induced from the original $\pi(\mathbf{w})$, and they are

(i) the market-type appearance probability function $\pi_M(\mathbf{w}) = \pi_2(\mathbf{w}) + \pi_4(\mathbf{w})$,
(ii) the limit-type appearance probability function $\pi_L(\mathbf{w}) = \pi_1(\mathbf{w}) + \pi_3(\mathbf{w})$,
(iii) the buy-type appearance probability $\pi_B(\mathbf{w}) = \pi_1(\mathbf{w}) + \pi_2(\mathbf{w})$,
(iv) the sell-type appearance probability $\pi_S(\mathbf{w}) = \pi_3(\mathbf{w}) + \pi_4(\mathbf{w})$.

Here, we have $\pi_M(\mathbf{w}) + \pi_L(\mathbf{w}) = 1$ and also $\pi_B(\mathbf{w}) + \pi_S(\mathbf{w}) = 1$ for all $\mathbf{w} \in W$.

Note that the value of $\pi_L(\mathbf{w})$ only depends on the spread of $\mathbf{w}$, and moreover, $\pi_L(\mathbf{w}) = f(s(\mathbf{w}))$, where $f : [\underline{s}, \overline{a}] \to [0,1]$ was defined in Section 2.3. Similarly, the value of $\pi_B(\mathbf{w})$ only depends on the mid-price of $\mathbf{w}$, and specifically, we let $\pi_B(\mathbf{w}) = h(m(\mathbf{w}))$, where $h : [\underline{s}/2, \overline{a} - \underline{s}/2] \to [0,1]$.

Recall that $f(s)$ is an increasing function of $\log s$, so $\pi_L(\mathbf{w})$ is increasing with $\log s(\mathbf{w})$, and $\pi_M(\mathbf{w})$ is decreasing with $\log s(\mathbf{w})$. Moreover, if $\mathbf{w}$ belongs to the region

$$W_M = \{\mathbf{w} : \underline{s} \le s(\mathbf{w}) < (1+\alpha)\underline{s}\},$$

we have $\pi_M(\mathbf{w}) = 1$ and $\pi_L(\mathbf{w}) = 0$. If $\mathbf{w}$ belongs to the region

$$W_L = \{\mathbf{w} : (1-\gamma)\overline{a} < s(\mathbf{w}) \le \overline{a}\},$$

we have $\pi_L(\mathbf{w}) = 1$ and $\pi_M(\mathbf{w}) = 0$. Here, $0 < \alpha < 1$, and $\alpha/(1+\alpha) \le \gamma < 1$.

Notice that $h(m)$ is a decreasing function of $\log m$, so $\pi_B(\mathbf{w})$ is decreasing with $\log m(\mathbf{w})$, and $\pi_S(\mathbf{w})$ is increasing with $\log m(\mathbf{w})$. Moreover, if $\mathbf{w}$ belongs to the region

$$W_B = \{\mathbf{w} : \underline{s}/2 \le m(\mathbf{w}) < (1+\delta)\underline{s}/2\},$$

we have $\pi_B(\mathbf{w}) = 1$ and $\pi_S(\mathbf{w}) = 0$. If $\mathbf{w}$ belongs to the region

$$W_S = \{\mathbf{w} : (1-\epsilon)(\overline{a} - \underline{s}/2) < m(\mathbf{w}) \le \overline{a} - \underline{s}/2\},$$

we have $\pi_S(\mathbf{w}) = 1$ and $\pi_B(\mathbf{w}) = 0$. Here, $0 < \delta < 1$, and $\delta/(1+\delta) \le \epsilon < 1$.

Definition 3.10. The *buffering region* of W is the largest nonclosed subset $H \subset W$ with the property that $\prod_{x \in \{L,M,B,S\}} \pi_x(\mathbf{w}) = 0$ for all $\mathbf{w} \in H$.

At any state $\mathbf{w} \in H$, there exists at least an $x \in \{L, M, B, S\}$ such that $\pi_x(\mathbf{w}) = 0$. Since $\pi_L + \pi_M = \pi_B + pi_S = 1$, there also exists at least a $y \in \{L, M, B, S\}$ such that $\pi_y(\mathbf{w}) = 1$ at the state $\mathbf{w}$. So we can have at most two elements, say, $x_1 \in \{L, M\}$ and $x_2 \in \{B, S\}$, such that $\pi_{x_1}(\mathbf{w}) = \pi_{x_2}(\mathbf{w}) = 0$, and $\pi_y(\mathbf{w}) = 1$ for $y \notin \{x_1, x_2\}$.

Definition 3.11. The *kernel region* of W is the largest closed subset $K \subseteq W$ with the property that $\pi_x(\mathbf{w}) \ne 0$ for all $x \in \{L, M, B, S\}$ and for all $\mathbf{w} \in K$.

Note that for all $\mathbf{w} \in K$ we also have $\pi_x(\mathbf{w}) \neq 1$ for all $x \in \{L, M, B, S\}$, as there exists a unique y such that $\pi_x(\mathbf{w}) = 1 - \pi_y(\mathbf{w})$, where $\pi_y(\mathbf{w}) \neq 0$.

In general, we have $K = W \setminus H$, and $K \cap H = \emptyset$. Thus we have a bipartition of the domain W, as $K \cup H = W$ and $K \cap H = \emptyset$. Since K is defined to be closed, and H is defined to be nonclosed, H may be empty, but $H \neq W$, and hence K is always nonempty. If $K = W$, then $H = \emptyset$. If $K = \{\mathbf{w}\}$, for some $\mathbf{w} \in W$ certain, then $H = W \setminus \{\mathbf{w}\}$.

In our framework, we have $\pi_L(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_M$, $\pi_M(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_L$, $\pi_B(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_S$, and $\pi_S(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_B$. Thus

$$H = W_L \cup W_M \cup W_B \cup W_S, \tag{9}$$

and $K = W \setminus H$, where $W_L \cap W_M = \emptyset$, and $W_B \cap W_S = \emptyset$ (see Fig. 6).

Note that K is closed, while H is not closed, but $H \cup \partial K$ is also closed, where ∂K denotes the boundary of the kernel region K.

Definition 3.12. The *s-range* of $R \subseteq W$ is

$$r_s(R) = \sup_{\mathbf{w}\in R} s(\mathbf{w}) - \inf_{\mathbf{w}\in R} s(\mathbf{w}). \tag{10}$$

Definition 3.13. The *m-range* of $R \subseteq W$ is

$$r_m(R) = \sup_{\mathbf{w}\in R} m(\mathbf{w}) - \inf_{\mathbf{w}\in R} m(\mathbf{w}). \tag{11}$$

Proposition 3.14. If the buffering region $H \neq \emptyset$, and the kernel region K satisfies

$$\min\{r_s(K), r_m(K)\} > \alpha(1+\alpha)(2+\alpha)\underline{s}, \tag{12}$$

the dynamical trading system (W, F, σ) is stochastically stable within K.

Proof. For any trajectory starting from a bid-ask pair $\mathbf{w} \in W$, all the possible states in W are finite, as shown in the proof of Lemma 3.7. The set of all the possible states for any initial state $\mathbf{w}$ can be denoted by a corresponding lattice $\Lambda(\mathbf{w})$. Let the neighborhood of any $\mathbf{v} \in \Lambda(\mathbf{w})$ be

$$N(\mathbf{v}) = \{\mathbf{v}S_1, \mathbf{v}S_2, \mathbf{v}S_3, \mathbf{v}S_4\} \cap W.$$

Observe that $\Lambda(\mathbf{w})$ is globally defined on $W = K \cup H$, so there exist states $\mathbf{v} \in \Lambda(\mathbf{w})$ near ∂K, such that $N(\mathbf{v}) \cap H \neq \emptyset$ and $N(\mathbf{v}) \cap K \neq \emptyset$.

Note that the spread $s(\mathbf{v})$ of the state $\mathbf{v}$ can be updated to $s(\mathbf{v})/(1+\alpha)$ or $(1+\alpha)s(\mathbf{v})$, and the mid-price $m(\mathbf{v})$ of the same state $\mathbf{v}$ can be updated to maximally $m(\mathbf{v}) + \alpha s(\mathbf{v})/2$ and minimally $m(\mathbf{v}) - \alpha s(\mathbf{v})/2$, so

$$r_s(N(\mathbf{v})) = \frac{\alpha(2+\alpha)}{1+\alpha} s(\mathbf{v}),$$
$$r_m(N(\mathbf{v})) = \alpha s(\mathbf{v}).$$

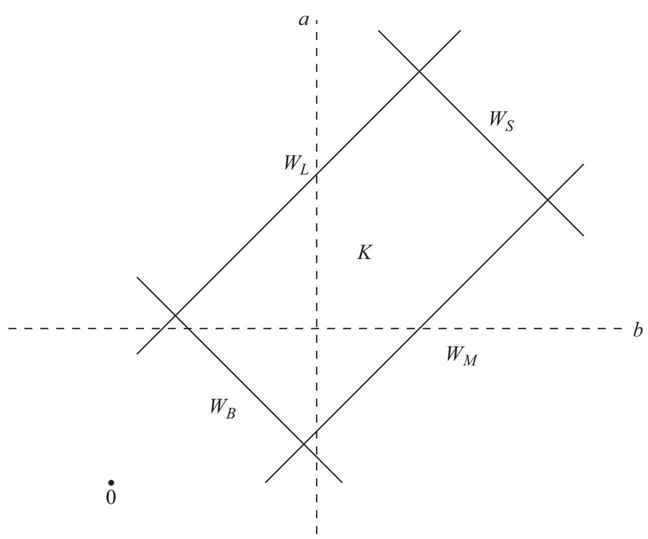

Fig. 6. The kernel region and buffering region.

Suppose $s(\mathbf{v})/(1+\alpha) \geq (1+\alpha)\underline{s}$, where $(1+\alpha)\underline{s}$ is the lower bound of the spread in K, so we have

$$r_s(N(\mathbf{v})) \geq \alpha(1+\alpha)(2+\alpha)\underline{s},$$
$$r_m(N(\mathbf{v})) \geq \alpha(1+\alpha)^2\underline{s},$$

and hence

$$r_s(K) > \inf_{\mathbf{v}\in K} r_s(N(\mathbf{v})),$$
$$r_m(K) > \inf_{\mathbf{v}\in K} r_m(N(\mathbf{v})).$$

So for any initial state $\mathbf{w} \in W$, there exists at least a $\mathbf{v} \in \Lambda(\mathbf{w}) \cap K$ such that $N(\mathbf{v}) \subset K$.

Note that $W = K \cup H$, and both K and H are nonempty, so both H and K are proper subsets of W. Suppose $\mathbf{w} \in H$. We know $H = W_L \cup W_M \cup W_B \cup W_S$, so there exists at least an $x \in \{L, M, B, S\}$ such that $\mathbf{w} \in W_x$. Note that $\pi_x(\mathbf{v}) = 1$ for all $\mathbf{v} \in W_x$, thus $\mathbf{w}$ will move towards K along a continuous flow in $\Lambda(\mathbf{w}) \cap H$. Since the number of states in $\Lambda(\mathbf{w}) \cap H$ is finite, $\mathbf{w}$ will move into the closed kernel region K after finite periods. If the market is stochastically stable within K once $\mathbf{w} \in K$, the market will function for infinite periods, and hence it is also stochastically stable with the initial state $\mathbf{w} \in H$. Thus we only need to show the trading system is stochastically stable for any initial state $\mathbf{w} \in K$. Let $p(\mathbf{w})$ denote the probability that bid-ask pairs stay in H with an initial state $\mathbf{w}$. We need to show that $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$.

Suppose $\mathbf{w} \in K$, then we have two possibilities, $N(\mathbf{w}) \subset K$, and $N(\mathbf{w}) \cap H \neq \emptyset$. If $N(\mathbf{w}) \cap H \neq \emptyset$, then $N(\mathbf{w}) \cap K \neq \emptyset$, otherwise $\mathbf{w} \in H$. There exist some $x \in \{L, M, B, S\}$, such that $\pi_x(\mathbf{v}) = 0$ for all $\mathbf{v} \in N(\mathbf{w}) \cap H$, and $\pi_x(\mathbf{v}) = \varepsilon_x \in [0, 1]$ for all $\mathbf{v} \in N(\mathbf{w}) \cap K$. $\pi_x(\mathbf{v}) \neq 1$ for all $\mathbf{v} \in K$ and all $x \in \{L, M, B, S\}$, otherwise there must exist a $y \in \{L, M, B, S\}$ such that $\pi_y(\mathbf{v}) = 1 - \pi_x(\mathbf{v}) = 0$, which then implies $\mathbf{v} \in H$, a contradiction. Hence $\varepsilon_x \neq 1$ for all $x \in \{L, M, B, S\}$.

If $\mathbf{w}$ moves to $\mathbf{v} \in N(\mathbf{w}) \cap H$ with a probability ε_x, it will return back to $\mathbf{v}' \in N(\mathbf{v}) \cap K$ with probability 1 in the next period, where $N(\mathbf{v}') \cap H \neq \emptyset$ as $\mathbf{v} \in N(\mathbf{v}')$. If $\mathbf{w}$ moves to $\mathbf{v} \in N(\mathbf{w}) \cap K$ with a probability $1 - \varepsilon_x$, it can stay within $\Lambda(\mathbf{w}) \cap K$ with $k(\mathbf{w})$ consecutive periods, and then move into a state $\mathbf{v}'$ such that $N(\mathbf{v}') \cap H \neq \emptyset$. Recall that, for all $\mathbf{w} \in K$, we have

$$\{\mathbf{v} \in \Lambda(\mathbf{w}) \cap K : N(\mathbf{v}) \subset K\} \neq \emptyset,$$

so $k(\mathbf{w}) \geq 0$. Then we have

$$p(\mathbf{w}) = \lim_{T\to\infty} \varepsilon_x p(\mathbf{w})\Big(1 - \frac{2}{T}\Big) + (1-\varepsilon_x)p(\mathbf{w})\Big(1 - \frac{k(\mathbf{w})+1}{T}\Big),$$

where $0 \leq k(\mathbf{w}) \leq T - 1$. When $k(\mathbf{w}) = T - 1$,

$$p(\mathbf{w}) = \lim_{T\to\infty} \varepsilon_x p(\mathbf{w})\Big(1 - \frac{2}{T}\Big) = \varepsilon_x p(\mathbf{w}),$$

which generates $(1-\varepsilon_x)p(\mathbf{w}) = 0$. Since $1 - \varepsilon_x \neq 0$, $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$ with $N(\mathbf{w}) \cap H \neq \emptyset$.

If $\mathbf{w} \in K$ and $N(\mathbf{w}) \subset K$, its trajectory can either achieve a state $\mathbf{w}' \in \Lambda(\mathbf{w}) \cap K$ such that $N(\mathbf{w}') \cap H \neq \emptyset$ after $h(\mathbf{w})$ periods, where $h(\mathbf{w}) \geq 1$, or never move to such a state $\mathbf{w}'$ and thus stay within K for ever. Note that $p(\mathbf{w}') = 0$, if $\mathbf{w}' \in K$ and $N(\mathbf{w}') \cap H \neq \emptyset$. So $p(\mathbf{w}) \leq p(\mathbf{w}') = 0$, but $p(\mathbf{w}) \geq 0$, hence $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$ with $N(\mathbf{w}) \subset K$.

As a result, $p(\mathbf{w}) = 0$ if $\mathbf{w} \in K$, and thus $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in W$. So the dynamical trading system is stable within K almost surely. □

Since (b_t, a_t) will stay within K almost surely, the random trajectories of b_t and a_t will also stay within bounded intervals. Note that

$$(1+\alpha)\underline{s} \leq s_t \leq r_s(K) + (1+\alpha)\underline{s}, \tag{13}$$

and

$$(1+\delta)\underline{s}/2 \leq m_t \leq r_m(K) + (1+\delta)\underline{s}/2, \tag{14}$$

almost surely for all $t \in \mathbb{Z}$, so we almost surely have

$$a_t \leq r_m(K) + r_s(K)/2 + (2+\alpha+\delta)\underline{s}/2 < r_m(K) + r_s(K)/2 + 2\underline{s}, \tag{15}$$

and

$$b_t \leq r_m(K) + (2+\alpha+\delta)\underline{s}/2 < r_m(K) + 2\underline{s}, \tag{16}$$

where $\alpha, \delta \in (0, 1)$, so $(2+\alpha+\delta)/2 < 2$.

The upper bounds of b_t and a_t have interesting implications on the roles of s-range and m-range in the limit order market. The upper bound of the best bid b_t in the market is solely determined by the m-range of the kernel region K, rather than any property of the whole domain W. The upper bound of the best ask a_t is also only related to the kernel region K, and yet it is determined by both the m-range and the s-range of K. Notice that the lower bounds of b_t and a_t are both close to $2\underline{s}$, so the bid-range in the market is approximately equal to $r_m(K)$, and the ask-range is roughly $r_m(K) + r_s(K)/2$. Evidently, the ask side of the limit order book should be more volatile than its bid side, in case that the range of a price is a meaningful indicator of its volatility.

4. Controlled Trading System

In this section, we assume again $H \neq \emptyset$, but either the s-range or the m-range of K will be less than $\alpha(1+\alpha)(2+\alpha)\underline{s}$. So the condition on the kernel region in Proposition 3.14 is no longer satisfied. We want to check whether a random trajectory of bid-ask pairs can maintain the property of stochastic stability within certain bounded domains.

Let $U_1 = W_L \cup W_M$. Since $K \neq \emptyset$, we have $W_L \cap W_M = \emptyset$ and $W \setminus U_1 \neq \emptyset$. Define $U_2 = W \setminus U_1$, so $W = U_1 \cup U_2$, and $U_1 \cap U_2 = \emptyset$.

Similarly, let $V_1 = W_B \cup W_S$, again $W \setminus V_1 \neq \emptyset$. Define $V_2 = W \setminus V_1$, so $W = V_1 \cup V_2$, and $V_1 \cap V_2 = \emptyset$.

Note that $K = U_2 \cap V_2$ and $H = U_1 \cup V_1$. Also observe that $r_s(K) = r_s(U_2)$ and $r_m(K) = r_m(V_2)$.

4.1. Controlled Spread Dynamics

By the condition $W_L \cap W_M = \emptyset$, we have $r_s(U_2) \geq 0$, so $(1+\alpha)\underline{s} \leq (1-\gamma)\overline{a}$, or

$$\underline{s}/\overline{a} \leq \frac{1-\gamma}{1+\alpha}.$$

At the same time, we assume that the s-range of U_2 is sufficiently small, namely, $r_s(U_2) < \alpha(1+\alpha)\underline{s}$, so

$$(1-\gamma)\overline{a} - (1+\alpha)\underline{s} < \alpha(1+\alpha)\underline{s},$$

which implies that $\underline{s}/\overline{a}$ has a lower bound,

$$\underline{s}/\overline{a} > \frac{1-\gamma}{(1+\alpha)^2}.$$

Intuitively, the upper bound of the s-range of U_2 gives such a sufficient condition that any type of marginal trader will definitely update any $\mathbf{w} \in U_2$ to some $\mathbf{w}' \in U_1$.

In brief, if $0 \leq r_s(U_2) < \alpha(1+\alpha)\underline{s}$, the domain W will have the following property,

$$\frac{1-\gamma}{(1+\alpha)^2} < \underline{s}/\overline{a} \leq \frac{1-\gamma}{1+\alpha}. \tag{17}$$

Once the above inequality is satisfied by the domain W, we have $U_2 \neq \emptyset$, and

$$U_2 = \{\mathbf{w} : (1+\alpha)\underline{s} \leq s(\mathbf{w}) \leq (1-\gamma)\overline{a}\},$$

where $s(\mathbf{w})$ is the spread function. Let the boundary of U_2 be ∂U_2, then

$$\partial U_2 = \{\mathbf{w} : s(\mathbf{w}) = (1+\alpha)\underline{s}\} \cup \{\mathbf{w} : s(\mathbf{w}) = (1-\gamma)\overline{a}\}.$$

Proposition 4.1. If the buffering region $H = U_1 \cup V_1$ is nonempty, and the kernel region $K = U_2 \cap V_2$ satisfies

$$0 \leq r_s(U_2) < \alpha(1+\alpha)\underline{s}, \tag{18}$$

$$r_m(V_2) > \alpha(1+\alpha)(2+\alpha)\underline{s}, \tag{19}$$

the dynamical trading system (W, F, σ) is stochastically stable within the region

$$\{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1+\alpha)^3\underline{s}\} \cap V_2.$$

Proof. Since $U_2 \neq \emptyset$ and its s-range is less than $\alpha(1+\alpha)\underline{s}$, we obtain

$$(1+\alpha)\underline{s} \leq (1-\gamma)\overline{a} < (1+\alpha)^2\underline{s}.$$

Note that

$$\min_{\mathbf{w}\in U_2} s(\mathbf{w}) = (1+\alpha)\underline{s},$$

$$\max_{\mathbf{w}\in U_2} s(\mathbf{w}) = (1-\gamma)\overline{a},$$

so we have

$$\underline{s} \leq \frac{(1-\gamma)\overline{a}}{(1+\alpha)} < (1+\alpha)\underline{s} = \min_{\mathbf{w}\in U_2} s(\mathbf{w}),$$

and also

$$(1+\alpha)^2\underline{s} > (1-\gamma)\overline{a} = \max_{\mathbf{w}\in U_2} s(\mathbf{w}).$$

Since any initial state $\mathbf{w} \in H$ will definitely move into K after finite periods, we only need to show the statement is true when the initial state $\mathbf{w} \in K$. Consider any initial state $\mathbf{w} \in U_2 \cap V_2$, we have $(1+\alpha)\underline{s} \leq s(\mathbf{w}) \leq (1-\gamma)\overline{a}$. Note that $\pi_L(\mathbf{w}) > 0$ and $\pi_M(\mathbf{w}) > 0$ for all $\mathbf{w} \in K$, so $\pi_i(\mathbf{w}) \neq 0$ for all $\mathbf{w} \in K$ and all $i \in \{1,2,3,4\}$. Suppose $\mathbf{w}$ is updated to $\mathbf{v} = \mathbf{w}S_1$ by a marginal trader of type $\sigma(1)$, $s(\mathbf{v}) = s(\mathbf{w})/(1+\alpha)$, which is greater than $\underline{s}$ and less than $(1-\gamma)\overline{a}/(1+\alpha)$. Since $(1-\gamma)\overline{a}/(1+\alpha) < \min_{\mathbf{w}\in U_2} s(\mathbf{w})$, we should have $\mathbf{v} \in W_M$, and hence $\pi_M(\mathbf{v}) = 1$. The trader in the next period will be market-type, namely, either type $\sigma(4)$ or type $\sigma(2)$. If she is of type $\sigma(4)$, $\mathbf{v}$ will then become $\mathbf{w}S_1S_4 = \mathbf{w}$, since $\{\sigma(1), \sigma(4)\}$ is a minimal periodic block. If she is of type $\sigma(2)$, $\mathbf{v}$ will become $\mathbf{v}' = \mathbf{w}S_1S_2$, where

$$S_1S_2 = \begin{pmatrix} \frac{1}{1+\alpha} & 0 \\ \frac{\alpha}{1+\alpha} & 1 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \\ 0 & 1+\alpha \end{pmatrix} = \frac{1}{1+\alpha} \begin{pmatrix} 1 & -\alpha \\ \alpha & 1+2\alpha \end{pmatrix},$$

so $s(\mathbf{v}') = s(\mathbf{w})$. Thus after two periods, $\mathbf{w}$ will return back to itself or move to a state with the same spread as itself.

If $\mathbf{w}$ is updated to $\mathbf{v} = \mathbf{w}S_2$ by a marginal trader of type $\sigma(2)$, then $\mathbf{v} \in W_L$. So in the next period, there will come either type-$\sigma(3)$ trader or type-$\sigma(1)$ trader. The type-$\sigma(3)$ trader will update $\mathbf{v}$ to $\mathbf{w}S_2S_3 = \mathbf{w}$, as $\{\sigma(2), \sigma(3)\}$ is a minimal periodic block. The type-$\sigma(1)$ trader will update $\mathbf{v}$ to $\mathbf{w}S_2S_1$ that has the same spread as $\mathbf{w}$. If $\mathbf{w}$ is updated to $\mathbf{w}S_3$ by a marginal trader of type $\sigma(3)$, $\mathbf{w}S_3$ will then be updated either to $\mathbf{w}S_3S_2 = \mathbf{w}$, or to $\mathbf{w}S_3S_4$ that have the same spread as $\mathbf{w}$. Finally, if $\mathbf{w}$ is updated to $\mathbf{w}S_4$ by a marginal trader of type $\sigma(4)$, $\mathbf{w}S_4$ will then move either to $\mathbf{w}S_4S_1 = \mathbf{w}$, or to $\mathbf{w}S_4S_3$ such that $s(\mathbf{w}S_4S_3) = s(\mathbf{w})$.

Therefore, any initial state $\mathbf{w} \in U_2 \cap V_2$ will be updated to a state with the same spread as itself after two consecutive periods. Note that the dynamics in (W, F, σ) is repeatedly composed of those two-period dynamical blocks, thus the spread that can be achieved in such a dynamical trading system will be greater than or equal to

$$\min_{\mathbf{w}\in U_2} s(\mathbf{w})/(1+\alpha) = \underline{s},$$

and less than or equal to

$$(1+\alpha) \max_{\mathbf{w}\in U_2} s(\mathbf{w}) < (1+\alpha)^3 \underline{s},$$

as $\max_{\mathbf{w}\in U_2} s(\mathbf{w}) < (1+\alpha)^2 \underline{s}$.

So the trajectory of bid-ask pairs starting from any initial state $\mathbf{w} \in W$ will be bounded in the region $\{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1+\alpha)^3 \underline{s}\} \cap V_2$ almost surely. □

If $0 \leq r_s(U_2) < \alpha(1+\alpha)\underline{s}$, there exists a unique exponent $l \in [1, 2)$, such that

$$\underline{s}/\overline{a} = \frac{1-\gamma}{(1+\alpha)^l}. \tag{20}$$

Note that the upper bound of the bid-ask spread in W is $\overline{a}$, so $(1+\alpha)^3\underline{s}$ should be no greater than $\overline{a}$, or equivalently,

$$\underline{s}/\overline{a} \leq \frac{1}{(1+\alpha)^3}.$$

Thus we need $(1-\gamma)(1+\alpha)^h \leq 1$, where $h = 3 - l$, so $h \in (1, 2]$. It appears to be a slightly stricter requirement than $(1-\gamma)(1+\alpha) \leq 1$ that we have used before, since $(1+\alpha)^h > 1+\alpha$ for $h > 1$.

Recall that $U_2 = \{\mathbf{w} : (1+\alpha)\underline{s} \leq s(\mathbf{w}) \leq (1-\gamma)\overline{a}\}$, we thus have two nonempty regions in U_1, which will contain buffering overflows,

$$\{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1+\alpha)\underline{s}\},$$

$$\{\mathbf{w} : (1-\gamma)\overline{a} < s(\mathbf{w}) < (1+\alpha)^3\underline{s}\}.$$

Evidently, the trajectory of bid-ask pairs will not stay exactly within the kernel region $K = U_2 \cap V_2$, but within K and parts of the buffering region $U_1 \cap V_2$ in H.

As we can see from the proof of Proposition 4.1, there are eight possible two-period dynamical blocks for all $\mathbf{w} \in U_2 \cap V_2$. Half of them are minimal periodic blocks, so the bid-ask pair driven by each of them will return back to $\mathbf{w}$. While the remaining ones will update $\mathbf{w}$ to $\mathbf{w}' \neq \mathbf{w}$, with $s(\mathbf{w}') = s(\mathbf{w})$ (see Fig. 7, where $s(\mathbf{w})$ is replaced simply by s).

If $\mathbf{w}' \neq \mathbf{w}$, the distance (in the b-a plane) between $\mathbf{w}$ and $\mathbf{w}'$ is

$$\frac{\sqrt{2}\alpha}{1+\alpha} s(\mathbf{w}), \text{ or } \sqrt{2}\alpha s(\mathbf{w}),$$

and the absolute difference between $m(\mathbf{w})$ and $m(\mathbf{w}')$ is

$$\frac{\alpha}{1+\alpha} s(\mathbf{w}), \text{ or } \alpha s(\mathbf{w}).$$

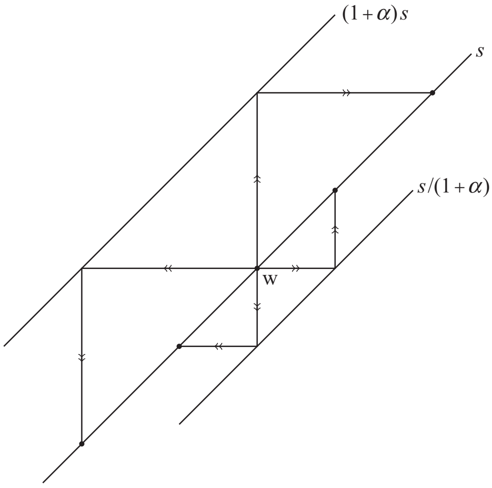

Fig. 7. Two-period nonperiodic blocks.

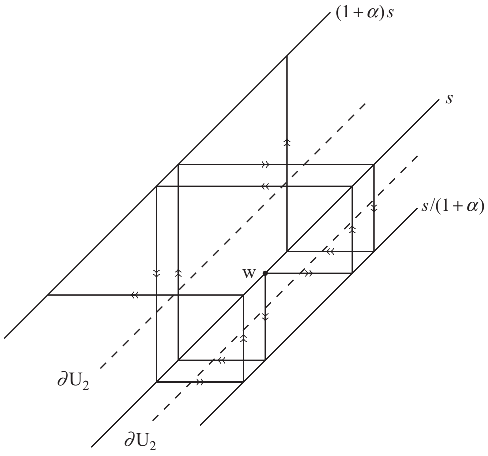

Fig. 8. Two trajectories starting from $\mathbf{w}$ in $U_2 \cap V_2$.

The possible trajectories in the dynamical trading system (W, F, σ) are consecutive combinations of those eight two-period dynamical blocks. Recall that $s(\mathbf{w}) \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$ for all $\mathbf{w} \in U_2 \cap V_2$. The trajectory starting from $\mathbf{w}$ will be bounded within the region

$$\{\mathbf{v} : s(\mathbf{w})/(1+\alpha) \leq s(\mathbf{v}) \leq (1+\alpha)s(\mathbf{w})\},$$

where

$$(1+\alpha)\underline{s} \in (s(\mathbf{w})/(1+\alpha), s(\mathbf{w})],$$
$$(1-\gamma)\overline{a} \in [s(\mathbf{w}), (1+\alpha)s(\mathbf{w})).$$

So the region containing buffering overflows are

$$\{\mathbf{v} : s(\mathbf{w})/(1+\alpha) \leq s(\mathbf{v}) < (1+\alpha)\underline{s}\},$$
$$\{\mathbf{v} : (1-\gamma)\overline{a} < s(\mathbf{v}) \leq (1+\alpha)s(\mathbf{w})\},$$

where $s(\mathbf{w})/(1+\alpha) \geq \underline{s}$, and $(1+\alpha)s(\mathbf{w}) < (1+\alpha)^3\underline{s}$. We show two trajectories starting from $\mathbf{w}$ with seven periods in the b-a plane (see Fig. 8, where $s = s(\mathbf{w})$).

4.2. Controlled Mid-Price Dynamics

If the m-range of V_2, rather than the s-range of U_2, is sufficiently small, we may establish a similar result to Proposition 4.1. Assume $0 \leq r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$. By the condition $r_m(V_2) \geq 0$, we have $V_2 \neq \emptyset$ and $W_B \cap W_S = \emptyset$, so $(1+\delta)\underline{s}/2 \leq (1-\epsilon)(\overline{a} - \underline{s}/2)$, or

$$\frac{\underline{s}/2}{\overline{a} - \underline{s}/2} \leq \frac{1-\epsilon}{1+\delta},$$

where $\delta \in (0,1)$ and $\epsilon \in [\delta/(1+\delta), 1)$. On the other hand, if $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$, then

$$(1-\epsilon)(\overline{a} - \underline{s}/2) - (1+\delta)\underline{s}/2 < \alpha(1+\alpha)\underline{s}/2,$$

which implies

$$\frac{\underline{s}/2}{\overline{a} - \underline{s}/2} > \frac{1-\epsilon}{1+\delta+\alpha(1+\alpha)}.$$

As a result, if $0 \leq r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$, the domain W will have the following property,

$$\frac{1-\epsilon}{1+\delta+\alpha(1+\alpha)} < \frac{\underline{s}/2}{\overline{a} - \underline{s}/2} \leq \frac{1-\epsilon}{1+\delta}. \qquad (21)$$

Under that condition, we have a nonempty V_2,

$$V_2 = \{\mathbf{w} : (1+\delta)\underline{s}/2 \leq m(\mathbf{w}) \leq (1-\epsilon)(\overline{a} - \underline{s}/2)\}.$$

Let the boundary of V_2 be ∂V_2, then

$$\partial V_2 = \{\mathbf{w} : m(\mathbf{w}) = (1+\delta)\underline{s}/2\} \cup \{\mathbf{w} : m(\mathbf{w}) = (1-\epsilon)(\overline{a} - \underline{s}/2)\}.$$

The lower bound in the condition (21) is a sufficient condition for the existence of dynamical two-period switching blocks, by which any state in $V_2 \cap U_2$ can return back into $V_2 \cap U_2$ after two periods. To confirm this claim, we only need to show that any state $\mathbf{w}$ in

$\partial V_2 \cap U_2$ at a time t will be updated to some $\mathbf{w}' \in V_1$ at the time $t+1$. Note that

$$\max_{\mathbf{v}\in V_2} m(\mathbf{v}) = (1-\epsilon)(\overline{a} - \underline{s}/2),$$

$$\min_{\mathbf{v}\in V_2} m(\mathbf{v}) = (1+\delta)\underline{s}/2,$$

and they can be achieved by the states in ∂V_2.

Let's consider a generic state $\mathbf{w} \in \partial V_2 \cap U_2$ with $m(\mathbf{w}) = \max_{\mathbf{v}\in V_2} m(\mathbf{v})$. A buy-type trader will induce a greater mid-price, which means that the next state will be in V_1. A sell-type trader can generate a new mid-price,

$$m(\mathbf{w}) - \frac{\alpha}{2}s(\mathbf{w}), \text{ or } m(\mathbf{w}) - \frac{\alpha}{2(1+\alpha)}s(\mathbf{w}).$$

A sufficient condition for the new state will be in V_1 is that the greatest mid-price should be less than $\min_{\mathbf{v}\in V_2} m(\mathbf{v})$, that's to say,

$$\max_{\mathbf{v}\in V_2} m(\mathbf{v}) - \frac{\alpha}{2(1+\alpha)} \min_{\mathbf{w}\in \partial V_2} s(\mathbf{w}) < \min_{\mathbf{v}\in V_2} m(\mathbf{v}).$$

Note that $s(\mathbf{w})/(1+\alpha) \geq (1+\alpha)\underline{s}$ for all $\mathbf{w} \in V_2 \cap U_2$, so $\min_{\mathbf{w}\in\partial V_2} s(\mathbf{w}) = (1+\alpha)^2\underline{s}$. Then we obtain

$$(1-\epsilon)(\overline{a} - \underline{s}/2) - \alpha(1+\alpha)\underline{s}/2 < (1+\delta)\underline{s}/2,$$

which is exactly equivalent to the condition $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$.

Now if $\mathbf{w} \in \partial V_2 \cap U_2$ with $m(\mathbf{w}) = \min_{\mathbf{v}\in V_2} m(\mathbf{v})$, a sell-type trader will induce a smaller mid-price, and hence the next state must be in V_1. While the new mid-price generated by a buy-type trader can be

$$m(\mathbf{w}) + \frac{\alpha}{2}s(\mathbf{w}), \text{ or } m(\mathbf{w}) + \frac{\alpha}{2(1+\alpha)}s(\mathbf{w}).$$

A sufficient condition ensuring the new state will be in V_1 is that the smallest mid-price should be greater than $\max_{\mathbf{v}\in V_2} m(\mathbf{v})$, that's to say,

$$\min_{\mathbf{v}\in V_2} m(\mathbf{v}) + \frac{\alpha}{2(1+\alpha)} \min_{\mathbf{w}\in \partial V_2} s(\mathbf{w}) > \max_{\mathbf{v}\in V_2} m(\mathbf{v}),$$

which is again equivalent to $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$.

Proposition 4.2. If the buffering region $H = V_1 \cup U_1$ is nonempty, and the kernel region $K = V_2 \cap U_2$ satisfies

$$0 \leq r_m(V_2) < \alpha(1+\alpha)\underline{s}/2, \tag{22}$$

$$r_s(U_2) > \alpha(1+\alpha)(2+\alpha)\underline{s}, \tag{23}$$

the dynamical trading system (W, F, σ) is stochastically stable within the region

$$\{\mathbf{w} : \underline{m} \leq m(\mathbf{w}) < \overline{m}\} \cap U_2,$$

where $\underline{m}$ and $\overline{m}$ are constants in the trading system.

The proof of Proposition 4.2 is roughly similar to that of Proposition 4.1, and thus not provided here. Once again, its trajectory of bid-ask pairs will not be bounded within the kernel region $K = V_2 \cap U_2$, but within K and parts of the buffering region $V_1 \cap U_2$, so that the trading system can contain certain buffering overflows to support its stability. In the limit, each state on the trajectory will be either in $K = V_2 \cap U_2$ or in the buffering region $V_1 \cap U_2$ with equal probability.

Since the updating process of the mid-price also depends on the spread of the states on the trajectory, the buffering region used to hold the buffering overflows is quite wide. We can see that the m-range of the region $\{\mathbf{w} : \underline{m} < m(\mathbf{w}) < \overline{m}\} \cap U_2$ is very large, by virtue of

$$\underline{m} = (1+\delta)\underline{s}/2 - \alpha(1-\gamma)\overline{a}/2, \tag{24}$$

$$\overline{m} = (1-\epsilon)(\overline{a} - \underline{s}/2) + \alpha(1-\gamma)\overline{a}/2, \tag{25}$$

where $(1-\gamma)\overline{a} = \max_{\mathbf{w}\in U_2} s(\mathbf{w})$. In fact, its m-range is

$$\overline{m} - \underline{m} = r_m(V_2) + \alpha(1-\gamma)\overline{a}, \tag{26}$$

which is greater than or equal to $\alpha(1-\gamma)\overline{a}$, and less than $\alpha(1-\gamma)\overline{a} + \alpha(1+\alpha)\underline{s}/2$. So $\overline{m} - \underline{m} = O(\overline{a})$. On the other hand, in Proposition 4.1, the s-range of the region

$$\{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1+\alpha)^3\underline{s}\} \cap V_2$$

is only $((1+\alpha)^3 - 1)\underline{s} = O(\underline{s})$, where $O(\underline{s}) \ll O(\overline{a})$.

Proposition 4.1 and 4.2 collectively produce interesting observations about the volatilities of the mid-price and the bid-ask spread on a limit order market. If the spread s in the kernel region $K = U_2 \cap V_2$ is required to be *ex ante* stable, then the random trajectory in the dynamical trading system will be bounded within a certain region with a sufficiently small s-range. However, if the mid-price m in K is required to be *ex ante* stable, the random trajectory can not remain in a certain region with a small m-range.

Suppose again the range of a time-dependent price can be thought of to be an indicator of its volatility. Let us adopt an artificial concept of “disorder” to describe the origins of price volatility, and yet we leave such an introduced concept in its obscurity. Moreover, imagine the disorder in a trading system could be classified into one intrinsic part and another external part (which are evidently similar to the terms of “self-generated disorder” and “quenched disorder” in physics), such that the intrinsic disorder has an *ex ante* controllable nature, while the external disorder has an *ex ante* uncontrollable nature, and hence carries a high volume of potential information in the trading system. In practice, more intrinsic disorder should make a price in the trading system less unpredictable, while more external disorder could make the price more unpredictable. And thus a price volatility that can be partially characterized by its unpredictability should be lower if the disorder is rather intrinsic, and higher if that is rather external.

Quite directly, we say that the disorder in the bid-ask spread is mainly intrinsic or self-generated, as we find that the bid-ask spread has an *ex ante* controllable nature. On the other hand, most of the disorder in the mid-price should be external or quenched, since we show that the mid-price has an *ex ante* uncontrollable nature. For this reason, we conclude that the volatility of the bid-ask spread should be lower than that of the mid-price in the dynamical trading system.

5. Final Remarks

In this work, we take a dynamical perspective to investigate the microstructure of the limit order market. A limit order market is theoretically considered as a dynamical trading system, in which sequential trading processes are determined by either deterministic or probabilistic switches between different types of traders (represented by their optimal trading decisions in a very short time interval). In our analysis, the perfect information usually required to model traders' strategic behavior is loosened to the knowledge of the so-called atomic trading schemes. We thus set less assumptions for traders and the market, and yet still obtain a powerful ability to understand and even predict the order flow and the order book evolution in a generic limit order market. Some interesting and hopefully insightful conclusions have developed, for instance, the best ask is more volatile than the best bid, the mid-price is more volatile than the spread, and the best ask seems to have more determinants than the best bid.

To close this work, we would like to make some concluding remarks on these two critical notions appearing in our main results — general instability in a deterministic trading system, and stochastic stability in a stochastic trading system.

The stability of a deterministic trading system means that there should exist some bid-ask pairs such that certain trading blocks can generate convergent limit cycles attracted to them. We show that a necessary condition for general stability of the trading system is the trading block must be reducible, so that they can generate periodic bid-ask pair dynamics. Thus the notion of general instability implies that there is no stable bid-ask pair in the trading system, and it happens if the trading block is irreducible.

Once the trading block in the dynamical trading system stochastically emerges from Σ_4^ω with a certain stationary probability measure, the trading system will not be unstable for sure. One intuitive reason is that a countably infinite random trading block will be reducible almost surely. Therefore, we state that the bid-ask pair dynamics will be bounded within the domain W for sure, which gives meanings to stochastic stability of the dynamical trading system. More strictly, we show that the domain that serves for stochastic stability is only a proper subset of W, which is defined as its kernel region. However, in the meantime, the buffering region of W still has a positive role for stochastic stability, as it occasionally holds states on the bid-ask pair trajectory.

Finally, it might be worth figuring out one practical application of the notions of kernel region and buffering region. They could be used to measure the risk of systemic instability in a real market. Say, the market would have a high risk of systemic instability, if its state moves into the buffering region, while the risk of systemic instability should be acceptable with a certain confidence level, if the state stays in the kernel region.

Acknowledgements

The author is grateful to P. Z. Maymin, P. Pellizzari, M. Tolotti, and the reviewers for their comments, either general or concrete.

References

Al-Suhaibani, M., Kryzanowski, L., 2000. An exploratory analysis of the order book, and order flow and execution on the Saudi Stock Market. J. Bank. Finance 24(8), 1323–1357.

Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. J. Finance 50(5), 1655–1689.

Demsetz, H., 1968. The cost of transacting. Q. J. Econ. 82(1), 33–53.

Farmer, J.D., Patelli, P., Zovko, I.I., 2005. The predictive power of zero intelligence in financial markets. Proc. Natl. Acad. Sci. USA 102(6), 2254–2259.

Foucault, T., Kadan, O., Kandel, E., 2005. Limit order book as a market for liquidity. Rev. Financ. Stud. 18(4), 1171–1217.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14(1), 71–100.

Harris, L., Hasbrouck, J., 1996. Market vs. limit orders: the SuperDOT evidence on order submission strategy. J. Financ. Quant. Anal. 31(2), 213–231.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53(6), 1315–1335.

Lehmann, B.N., Modest, D.M., 1994. Trading and liquidity on the Tokyo Stock Exchange: A bird's eye view. J. Finance 49 (3), 951–984.

Parlour, C.A., 1998. Price dynamics in limit order markets. Rev. Financ. Stud. 11(4), 789–816.

Smith, E., Farmer, J.D., Gillemot, L., Krishnamurthy, S., 2003. Statistical theory of the continuous double auction. Quant. Finance 3(6), 481–514.

Sornette, D., 2003. Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press, Princeton, NJ.