Matching Algorithms of International Exchanges

Karel Janeček, Martin Kabrhel

December 1, 2007

Abstract. The aim of this paper is to analyze different kinds of trade matching algorithms. The matching (or trade allocation) algorithm is an important part of an exchange trading mechanism. We begin with an overview of the matching algorithms currently used at the biggest world derivatives exchanges. Then we analyze the impact of these algorithms on the strategy of a rational trader, and derive implications of the induced trader's behavior for the overall market efficiency. Our special focus is on the Time Pro-Rata algorithm introduced by Euronext.LIFFE in 2007 for the short-term interest rate futures contracts. Using rigorous mathematical models, we discuss how the optimal trading strategy should look like, and point out a number of unusual properties of this strategy. The obtained results might be interesting not only from the theoretical point of view, but also for a practical trader. Our analysis implies that the Time Pro-Rata algorithm substantially complicates decision making, and, more importantly, induces individually rational trader's behavior that is inconsistent with the general market efficiency.

Keywords: *Pro-Rata algorithm, Price/Time algorithm, international derivative exchanges, efficiency of financial markets, Time Pro-Rata algorithm*

JEL Classification: G15

1 Motivation

Financial markets provide an essential tool for an efficient functioning of today's global economy. An important part of financial markets are derivative exchanges. By trading derivatives, economic agents are able to mitigate or even remove the financial components of the risks associated with their primary businesses. In order to be able to effectively use financial markets the hedgers need liquidity and, if possible, low trading costs.

Efficiency and liquidity of a financial market strongly depends on the trading mechanism. Nowadays, most markets are electronic and trading is done directly from a computer screen or automatically by so-called APIM (automated price injection) models. In order to attract liquidity, an electronic exchange has a need for so-called *market makers* or *price*

[0] RSJ Invest, a.s., e-mails: karel.janecek@rsj.cz; martin.kabrhel@rsj.cz

makers, i.e., agents who place limit orders on the bid and/or offer sides. A hedger or other market participants (so called *price takers*) are able to make a direct trade for the best price offered by price makers.

The motivation of market makers to place limit orders is to “buy low and sell high”. Thus, a market maker would prefer as large spread between his bid and offer as possible. However, the competition among market makers usually tends to narrow the spread to a minimum size of one *tick*. With a relatively narrow spread, not all trades done by market makers are profitable. In order to be successful, a market maker must be able to offset the losing (bad) trades by a sufficient number of profitable (good) trades so that, in expected value, an upcoming random trade is profitable.

The size of one tick is determined by an exchange. The exchange’s goal is maximization of the volume traded. Setting too large tick size would discourage price takers from trading as the trading is too costly. On the other hand, no minimum spread would discourage price makers: When placing a limit order for a certain price, a competition could place a limit order for an arbitrarily better price. Similarly, the price makers would be discouraged if the minimum tick size was too low.

The existence of price makers is especially important for the exchange when a new market is developed. The exchange often chooses so called *designated market makers* who are required to place limit order of a minimum size and certain maximum spread (say 2 ticks). On the other hand, the designated market makers can usually trade for free and might also get financial compensation for their service.

Furthermore, for successful trading of price makers, the essential feature is the matching algorithm.

2 Matching Algorithms

In this section we describe the basic *matching* (or *trade allocation*) algorithms. We compare the advantages and disadvantages of each algorithm. We list matching algorithms for a few important derivative exchanges.

The trade allocation process. Consider an incoming market order of N lots. (By lot we mean the minimum traded volume.) Suppose that this market order is traded against n limit orders for the best price of a cumulative volume Q lots. A matching algorithm decides how many lots each limit order gets. The most common are the *Pro-Rata* and *Price/Time* matching algorithms.

2.1 Pro-Rata Algorithm

Under this algorithm, an incoming trade is split among limit orders proportionally to their sizes. More precisely: Suppose n limit orders $Q_1, ..., Q_n$ of the cumulative volume Q lots, and an incoming trade of $N \leq Q$ lots. Each order Q_i has the pro-rata proportion

$$p_i = \frac{Q_i}{Q}.$$

The number of lots obtained by order Q_i is equal to

$$P_i = \text{floor}(p_i N).$$

The remaining lots may be redistributed in various ways. Most common methods are:

- FIFO (first in, first out). This is a common solution that gives an extra bonus to the first order in the queue. (See Price/Time algorithm.)
- Unfilled orders get 1 lot. The remaining lots are assigned to those orders which pro-rata proportions are less than one. (In the order from the biggest to the smallest proportion.) This feature provides an extra motivation for small market participants.
- Analogues to the D'Hondt method. This is a method often used in the election process which operates with "demand quotients" (see [1]).

A frequent element of the pro-rata algorithm is time priority for the first price-improving limit order, usually up to a certain certain maximum number of lots. This feature motivates narrowing the spread between bid and offer.

An interesting question arises what would be in some sense "optimal" redistribution of the remaining lots. The D'Hondt's method is mathematically appealing, however, it is not commonly used in practice. One could argue that even better way of distributing the remaining lots would be by generating random numbers. The likelihood of obtaining an additional lot would be equal to the remaining decimal proportion for each market participant. For example, suppose there are two working limit orders of sizes 10 lots and 20 lots. Consider an incoming trade of 2 lots. Then the first order has the proportion of $\frac{1}{3} \cdot 2 = \frac{2}{3}$, while the second order has the proportion of $\frac{2}{3} \cdot 2 = \frac{4}{3}$. Thus, the second order receives 1 lot, and the remaining 1 lot is assigned to the first order with probability $\frac{2}{3}$, and to the second one with probability $\frac{1}{3}$. In expected value both traders obtain their precise proportions.

2.2 Price/Time Algorithm

Under the Price/Time algorithm, an incoming trade is matched with the the first orders (at the best price), which is the so called FIFO method.

More precisely: Suppose n limit orders $Q_1, ..., Q_n$ of a cumulative volume Q lots, and an incoming trade of $N \leq Q$ lots. Then there exists an index $1 \leq j \leq n$ such that $\sum_{i=1}^{j} Q_i \leq N$ and $\sum_{i=1}^{\min(j+1,n)} Q_i \geq N$. Under the Price/Time algorithm, the first j limit orders are filled in full, and the remaining lots are assigned to the $(j+1)$-th limit order (if $j < n$).

2.3 Comparison of Price/Time and Pro-Rata

Following are few basic remarks about the two basic algorithms and their comparison.

Price/Time algorithm

- Motivates to narrow the spread, since by narrowing the spread the limit order is the first in the order queue.

- Discourages other orders to join the queue since a limit order that joins the queue is the last.
- Might be computationally more demanding than pro-rata. The reason is that market participants might want to place more small orders in different positions in the order queue, and also tend to "flood" the market, i.e., place limit order in the depth of the market in order to stay in the queue.

An implication of the first two features is that the spread for Price/Time algorithm might tend to be rather narrow, while the cumulative quoted volume at a given price might be smaller than under the Pro-Rata algorithm.

Pro-Rata algorithm

- Motivates other orders to join the queue with large limit orders. As a consequence, the cumulative quoted volume at the best price is relatively large.
- Does not motivate to narrow the spread in the natural way. This weakness is partially offset by introducing the time priority element for the first order that makes a new price.

We note that the most common algorithm is Price/Time. We note that Pro-Rata is often used for derivative contracts traded with many expirations (eg. STIR futures).

2.4 International Derivative Exchanges

We now provide a summary of the most important international derivative exchanges. This summary is for general overview only and does not pretend to be complete.

- Euronext.LIFFE (London International Financial Futures Exchange).
  - Interest rate futures – Euribor (Euro Interbank Offered Rate), Short Sterling, Euroswiss.
    - Until middle 2007: Pro-rata algorithm.
    - Since middle 2007: Time pro-rata algorithm (see the next section).
- CME (Chicago Mercantile Exchange)
  - Eurodollar – pro-rata algorithm with FIFO for remaining lots.
- CBOT (Chicago Board of Trade)
  - Fed Funds Futures – pro-rata algorithm with preference for unfilled orders.
- Eurex (Frankfurt) – price/time algorithm.

3 LIFFE Time Pro-Rata Algorithm

In this section we provide theoretical analysis of the *Time Pro-Rata* algorithm which was introduced by one of the most important derivative exchanges Euronext.LIFFE in London in middle 2007. The results are somewhat surprising and very interesting. We next discuss the impact of the new algorithm on practical trading.

3.1 Time Pro-Rata Algorithm

Suppose n limit orders $Q_1, ..., Q_n$ of a cumulative volume Q lots, and an incoming trade of $N \leq Q$ lots. Each order Q_i has an order rank $k_i = i$. The first order in the queue is Q_n with order rank $k_n = n$, i.e. order ranks are integers from 1 to n. The time pro-rata proportion for limit order Q_i is given by

$$p_i = \frac{Q_i k_i}{\sum_{j=1}^{n} Q_j k_j}.$$

The number of lots obtained by order Q_i is equal to

$$P_i = \min\{Q_i, \text{floor}(p_i N)\}.$$

Unfilled orders might get 1 lot. Any remaining lots are distributed in additional rounds. The matching algorithm further includes the time priority element for the first price-improving order of a size 50 to 500 lots. For the original definition of the time pro-rata proportion see [2]. The Exchange definition is equivalent to ours, albeit uses somewhat cumbersome notation.

3.2 Rational Trading Under The Time Pro-Rata Algorithm

In the following we list the immediate consequences of the definition of the Time Pro-Rata algorithm, and then analyze the rational trading strategies.

After exploring the definition one can immediately see that there is a motivation for a market participant to split his single limit order to more orders. The reason is that the pro-rata proportion of a split order does not change, while the cumulative order rank can be increased. We can conclude:

- If we consider the first market participant, it can be easily seen that the first market participant will split his limit order of size L lots into L orders of minimal sizes 1 lot. This way any new incoming market participant will obtain a lower time pro-rata proportion, regardless of how large limit orders he places, whether splitting or not. (Intuitively, by splitting orders the first market participant "shades" any potential followers.)

- Regardless of the structure of an order book, a new market participant will split his order of size L lots into one order of size $L_1 \geq 1$ lots and $L - L_1$ orders of sizes 1 lot. Especially, if $L_1 = L$ then there is no splitting, while $L_1 = 1$ means full splitting like in the case of the first market participant. (If the agent placed a larger order to a latter place then by moving the extra lots to the first placed order with the highest order rank he would obviously improve his time pro-rata proportion.)

In the following example one can see that the splitting of orders is essential for the resulting traded proportion.

Example. Suppose an order book with five market participants. The first two and the last two market participants place one limit order of the size 120 lots. The third agent places orders of the total size 120 lots while using different trading strategies.

Consider an incoming trade of the size 100 lots. We compare the following four trading strategies: No splitting strategy, uniform splitting to three orders strategy, smart splitting to three orders strategy, and optimal splitting strategy.

No splitting strategy

Order number	Order size	Resulting fill
1	120	34
2	120	27
3	**120**	**20**
4	120	13
5	120	6

The fill for the no splitting strategy is 20 lots.

Uniform splitting to three orders strategy

Order number	Order size	Resulting fill
1	120	36
2	120	30
3	**40**	**8**
4	**40**	**6**
5	**40**	**5**
6	120	10
7	120	5

The combined fill for the uniform splitting strategy is 8+6+5 = 19 lots. Note that the uniform splitting results in lower total fill than the no splitting strategy. This is solely due to the rounding effect.

Smart splitting to three orders strategy

Order number	Order size	Resulting fill
1	120	34
2	120	28
3	**118**	**23**
4	**1**	**1**
5	**1**	**1**
6	120	9
7	120	4

The combined fill for this smart splitting strategy is 23+1+1 = 25 lots. This is an essentially larger fill than for the no splitting strategy.

Optimal splitting strategy

Order number	Order size	Resulting fill
1	120	34
2	120	32
3	**99**	**26**
4	**1**	**1**
5	**1**	**1**
...	...	...
8	**1**	**1**
9	**1**	**0**
...	...	...
24	**1**	**0**
25	120	2
26	120	1

The combined fill for the optimal splitting strategy is $26 + 5 \cdot 1 = 31$ lots. For the optimal splitting, the difference from the no splitting strategy is enormous. (The optimal splitting strategy is not easy to find.)

In the following theorem we show why the most "natural" strategy of placing a single large limit order is not a good one.

Theorem 1. *Let $n, L, Q_1, Q_2, ..., Q_n$ be positive integers. Assume we have n working limit orders of sizes $Q_1, Q_2, ..., Q_n$ lots. Suppose no remaining time priority order. Suppose further that there is no rounding, i.e. one can trade a fraction of a lot.*
The agent decides to place limit orders of a total size L lots such that the sequence of orders is non-increasing. Then the worst possible strategy is to place one limit order of the size L.

Proof. Suppose first that the agent places one limit order of size L. Then the matching proportion of $i-$th order is

$$p_i = \frac{(n+2-i)Q_i}{\sum_{j=1}^{n}(n+2-j)Q_j + L}, \quad i = 1, ..., n.$$

The agent has the matching proportion of

$$s_1 = \frac{L}{\sum_{j=1}^{n}(n+2-j)Q_j + L}.$$

Let k be a positive integer. Suppose now that the agent splits his order into k orders of sizes $L_1, \cdots L_k$, where $\sum_{j=1}^{k} L_j = L$ and $L_1 \geq L_2 \geq \ldots \geq L_k$. The agent's proportion changes to

$$s_2 = \frac{\sum_{j=1}^{k}(k+1-j)L_j}{\sum_{j=1}^{n}(n+1+k-j)Q_j + \sum_{j=1}^{k}(k+1-j)L_j}.$$

Denote $S = \sum_{j=1}^{k}(k+1-j)L_j$, $\alpha = \sum_{j=1}^{n}(n+1-j)Q_j$ and $\beta = \sum_{j=1}^{n} Q_j$. We want to prove the inequality $s_2 \geq s_1$ which is equivalent to

$$S(\alpha + \beta + L) \geq L(\alpha + k\beta + S),$$

equivalently

$$\alpha(S - L) \geq \beta(kL - S).$$

This holds because $\alpha \geq \beta$ (which is trivial), and $S - L \geq kL - S$ from the definition of S and from the fact that L_k is a non-increasing sequence. □

Remark 1. We make the following observations:

- We obtain equality $s_2 = s_1$ in Theorem 1 (i.e., splitting does not help) if and only if both $n = 1$ (only 1 preceding order) and we split equally, i.e., $L_1 = \ldots = L_k$.
- From remarks and Theorem 1 it follows that, for the optimal trading strategy, $L_2 = \ldots = L_k = 1$ and $k > 1$ (i.e., we split) as soon as $L > 1$.

3.3 The Optimal Trading Strategy for New Market Participant

We assume that a new market participant places a limit order. We first assume that the agent maximizes his matching proportion locally, i.e., without considering additional market participants that might join the order book. In Remark 3 we will provide an explicit formula for the case when the agent assumes additional fixed number of market participants.

We approximate the discrete order book and respective ranks by a continuous time version which enables us to obtain the explicit results.

Theorem 2. *Suppose that the order book consists of Q orders of the size 1 lot. The agent wants to place a limit order of the size L lots. The optimal strategy is to split a size pL into minimum 1 lot orders, and the remainder $(1-p)L$ lots are submitted to the market as the first order of the agent.The optimal proportion p is equal to*

$$p = p(z) = \frac{1}{2}\frac{\sqrt{1+4z}\; - 1}{z},$$

where $z = L/Q$ is the size of the agent's order relative to the existing orders in the market.

Proof. The weight of the agent is given as

$$w(p) = \int_0^{pL} u\,du + pL\cdot(1-p)L = \left(p - \frac{1}{2}p^2\right) L^2.$$

The matching proportion of the agent is given by

$$s(p) = \frac{w(p)}{w(p) + \int_{pL}^{pL+Q} u\,du} = \frac{w(p)}{w(p) + pLQ + \frac{1}{2}Q^2}.$$

We maximize the matching proportion by setting $s'(p) = 0$, and get

$$w'(p)\left(pLQ + \frac{1}{2}Q^2\right) = w(p)LQ,$$

equivalently

$$p^2 + p\cdot\frac{Q}{L} - \frac{Q}{L} = 0.$$

The positive solution of the previous equation is given by

$$p(z) = \frac{1}{2}\frac{\sqrt{1+4z}\; - 1}{z},$$

with $z = L/Q$. □

Remark 2. We can see that $p(z)$ is an decreasing function of $z \geq 0$. We have

$$\lim_{z\to 0_+} p(z) = 1, \qquad \lim_{z\to\infty} p(z) = 0.$$

This means that we split fully if the number of preceding one lot orders is large, while we do not split in the case where our agent has essentially larger order than the number of preceding one lot orders. For a realistic scenario of $z = 1$ we have $p(1) = \frac{1}{2}\left(\sqrt{5}-1\right) \doteq 0.618$ (the famous golden ratio, see [3]), i.e., the agent splits 61.8% of the desired limit order and the remaining 38.2% is submitted as the first large order.

Remark 3. Note that the agent might split more if he were to consider new market participants that would join the order book.

We can obtain explicit results in the case where the agent assumes that a fixed number of one lot limit orders will follow after he places his orders. Denote Q_u the number of one lot orders preceding the agent, L the agent's desired limit order with the optimal split proportion p, and Q_d the remaining one lot orders in the queue following the agent. Then

$$p(z_d, z_u) = \frac{1}{2}\left(\sqrt{\left(2z_d + \frac{z_d^2 + z_u^2}{z_u}\right)^2 + 4\frac{z_d^2 + z_u^2}{z_u}} - \left(2z_d + \frac{z_d^2 + z_u^2}{z_u}\right)\right),$$

where $z_d = Q_d/L$, $z_u = Q_u/L$.

By setting $z_d = 0$ we obtain as a special case Theorem 2. We can see that

$$\lim_{z_d \to \infty} p(z_d, z_u) = 1, \text{ and } \lim_{z_u \to \infty} p(z_d, z_u) = 1.$$

This means that we split split fully if either z_d or z_u are sufficiently large. For $z_d = z_u$ we have

$$p(z, z) = \frac{1}{2}\left(\sqrt{(4z)^2 + 8z} - 4z\right).$$

For example, if $z_d = z_u = 1$ we get $p(1, 1) = \left(\sqrt{6} - 2\right) \doteq 0.4495$, so in this case the agent splits only 44.95% of the limit order.

It is interesting to analyze the function $p(z_d, z_u)$ as a function of z_d for fixed z_u. We can see that $p(z_d, z_u)$ is firstly a decreasing function of z_d, reaches a minimum, and then $p(z_d, z_u)$ converges to 1 as $z_d \to \infty$. The intuition is that the following one lot orders first help the agent in a sense of "pushing" his first order forward, however, later the number of following one lot orders becomes significant and hurts the agent.

Remark 4. An interesting question arises concerning the resulting market equilibrium under the Time Pro-Rata algorithm with rational trading strategies used by all market participants. Under the assumption of optimal behavior of all market participants the order book would consist of mostly one lot orders with rarely appearing large orders. The first trader splits fully, while the second trader might place a large first order followed by minimal 1 lot orders. The size of the first order is a function of the intensity of coming market trades relative to the intensity of new limit trades. An analogous strategy applies for additional traders – always placing one order of possibly large size followed by a group of minimal 1 lot orders.

3.4 Rules And Regulations

Based on the theoretical analysis it appears that the optimal trading strategy involves a large amount of order splitting. In fact, in order to obtain a good position in the order book the market participants might be motivated to place limit orders also in the depth of the market, not only at the best price, similarly as in the case of Price/Time algorithm. Thus, each market participant might have an enormous number of limit orders. A large number of orders could cause even technical problems for the Exchange.

The negative consequence of the algorithm was later (after consultations with experts of Czech brokerage house RSJ Invest[1]) resolved by the Exchange by implementing several dis-

[1] References are available by the authors upon request.

cretionary rules that significantly impact the practical trading. One of the most important rules is the prohibition of order splitting. The Exchange explicitly states (see [2], paragraph 4.3): *"For the avoidance of doubt, members are prohibited from entering multiple orders at the same price level for the specific purpose of attempting to obtain a greater proportion of any subsequent matching volume than would be the case if a single order had been submitted (i.e. order splitting)."*

The discretionary feature of the previous rule is especially striking when one considers common practical trading behavior. A natural component of practical trading are order amends (increasing or decreasing the limit order size.) However, the Time Pro-Rata algorithm in principle prohibits amending orders up as the amended limit order would lose its position in the order book and go to the end of the queue. Thus, instead of amending an existing limit order up the market participant needs to place a new limit order. Note, however, that placing an additional limit order might or might not be considered as breaching the rule of "no order splitting".

Another somewhat striking rule is limiting the number of messages (i.e., number of limit trade placements and cancelations) per Exchange member. This rule was introduced after few members, including RSJ Invest, started to offer a very profitable trading strategy of small limit orders for clients. The Exchange decided to implement the messaging constraint per member (respectively per so-called *individual trading mnemonics*) rather than per client in order to limit the number of small clients per Exchange member. In principle, this rule implies that certain types of clients (the small clients) are penalized. The number of small clients per Exchange member is limited unless the Exchange member trades also for clients with large limit orders.

From the theoretical analysis we know that the motivation of a big player is to split his limit trade into many small limit trades which is prohibited by the "no order splitting" rule. Since a big trader is forced to trade in a detrimental way one can easily infer that the small market participants placing small orders (e.g. one lot orders) have a serious advantage as they trade "optimally".

4 Conclusion

The trading mechanism on today's electronic exchanges is an important component that has a great impact on the efficiency and liquidity of financial markets. The choice of matching algorithm is an important part of the trading mechanism. The most common matching algorithms are the Pro-Rata and Price/Time algorithms, with minor modifications related mostly to distributing the remaining lots after rounding, and, in the case of the Pro-Rata algorithm, to the time priority for the first price-improving limit order. Both algorithms have certain advantages and disadvantages, and both are used in practice, with Price/Time being more common.

In summer 2007 one of the major derivative exchanges Euronext.LIFFE introduced a modification called Time Pro-Rata algorithm. This matching algorithm exhibits a mathematical challenge for rigorous analysis, including exploration of the optimal trading strategies for market makers. We show that the optimal trading strategy implies enormous splitting of an intended limit order.

Unfortunately, following the optimal trading strategy may cause practical problems to the Exchange, even from the computational point of view of the central trading machine. This might be one of the reasons why the Exchange implements various discretionary regulations such as the prohibition of order splitting that turns out to be somewhat contradicting the natural trading behavior.

References

[1] D'Hondt method, *http://en.wikipedia.org/wiki/D'Hondt_method*

[2] LONDON NOTICE No. 2908, *http://www.euronext.com/fic/000/022/641/226413.pdf*

[3] The golden ratio, *http://mathworld.wolfram.com/GoldenRatio.html*