

Mixed FIFO/Pro Rata Match Algorithms

Robert Almgren and Eugene Krel

May 31, 2013

The NYSE LIFFE exchange has announced a change, effective May 29, 2013, to the pro rata trade matching algorithm for three month Short Sterling and Euroswiss futures contracts [LIFFE, 2013]. (*Update May 30:* LIFFE has announced that they are rolling back the change.) This note summarizes the content of this change, and compares it to the CME mixed time priority / pro rata match algorithm. We suggest that the CME algorithm is more effective at maintaining elements of both time priority and pro rata in the same match algorithm.

Suppose n orders are resting in the book at the best inside price, of sizes V_1, \dots, V_n in order of time submission. Let $V = V_1 + \dots + V_n$ be the total inside quote size. It is also useful to denote $P_j = \sum_{i=1}^j V_i$ for the total volume preceding and including order j . We have $P_0 = 0$, $P_j = P_{j-1} + V_j$ for $j = 1, \dots, n$, and $P_n = V$.

Suppose a market order of size Q arrives. If $Q \geq V$, then each resting limit order is completely filled. But if $Q < V$, then the incoming volume must be allocated among the n resting orders in some way that accomodates the goals of the exchange and of the market participants. The allocation method is called a *match algorithm*.

A wide variety of match algorithms may be expressed in the form

$$Q_j = \min\{V_j, f_j Q\}$$

where Q_j is the allocation given to the j th order and f_1, \dots, f_n are *allocation factors* with $f_1 + \dots + f_n = 1$. The quantities $f_j Q$ must be rounded to integers but we neglect that subtlety, as well as the special priority often given to the first order.

If each $f_j Q \leq V_j$ then $Q_j = f_j Q$ and $\sum Q_j = Q$; the incoming market order is fully allocated. But if any $f_j Q > V_j$, then $\sum Q_j < Q$. In that case, the formula is to be repeated with the orders that have not been completely filled.

Pro rata matching sets $f_j = V_j/V$. Time of submission is ignored, there is no penalty for cancelling and resubmitting limit orders, and quote volume tends to be very volatile. Participants must oversize to receive any fills [Field and Large, 2008].

Time priority or *FIFO* (first-in, first-out) sets $f_1 = 1$ and $f_j = 0$ for $j = 2, \dots, n$. If $Q > V_1$ then the order is not fully allocated on the first pass, and the algorithm is repeated to fill successive limit orders. Large early limit orders can completely block later entrants from receiving any fill volume.

Pro rata algorithms are widely used for short-term interest rate (STIR) products, where large tick and low volatility lead to large order queues. To improve market performance, exchanges experiment with various interpolations between pro rata



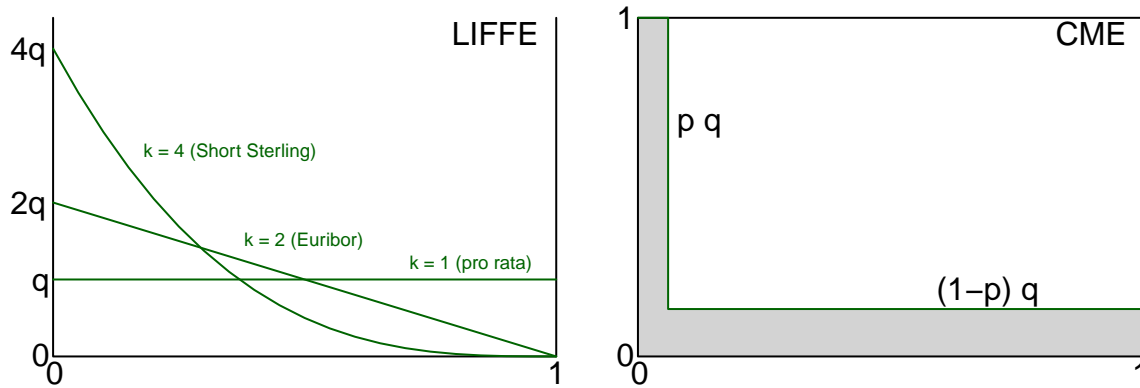


Figure 1: Mixed allocation algorithms, for small market order sizes; q denotes the size of the incoming market order as a fraction of total resting size. Horizontal axis is the position of a limit order in the queue as a fraction of total resting volume: 0 denotes the earliest order, 1 the last order entered. Vertical axis is the fraction of each order that is filled, neglecting rounding and assuming that q is small.

and FIFO algorithms, giving priority to early orders while still allowing late entrants to trade; McPartland [2013] and others suggest that this should be universal.

LIFFE matching algorithm

For its STIR products, LIFFE takes

$$f_j = \frac{(V - P_{j-1})^k - (V - P_j)^k}{V^k}.$$

Pro rata matching is obtained for $k = 1$, and FIFO in the limit $k \rightarrow \infty$. Intermediate values of k interpolate between the two. LIFFE sets $k = 2$ for Euribor, and $k = 4$ for Short Sterling and Euroswiss (effective May 29, 2013, [LIFFE, 2013]).

To understand this algorithm, note that by the Mean Value Theorem of calculus,

$$f_j = \frac{V_j}{V} g(x_j)$$

where

$$g(x) = \frac{d}{dx}(1-x)^k = k(1-x)^{k-1} \quad \text{and} \quad \frac{P_{j-1}}{V} < x_j < \frac{P_j}{V}.$$

If all limit orders are small fractions of the total, then $x_j \approx P_{j-1}/V \approx P_j/V$ and the fraction of each limit order that is filled depends only on its position in the queue. The fraction of order j that is filled is

$$\frac{Q_j}{V_j} = \min\{1, q g(x_j)\}.$$

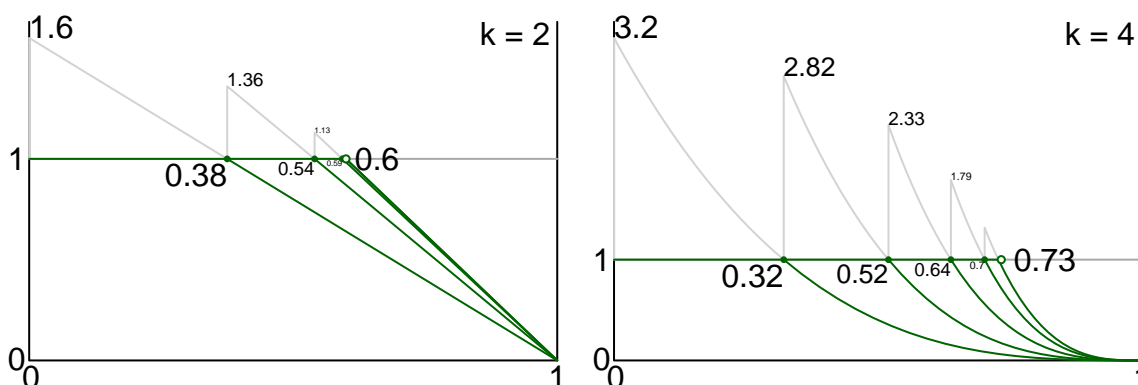


Figure 2: Convergence of LIFFE algorithm for large market orders. Left panel has $k = 2$ as for Euribor; right panel has $k = 4$ as for Short Sterling. As in Figure 1, the vertical axis is the fraction of a resting limit order that is filled, for a position in the queue given by the horizontal axis. The incoming market order has the very large size $q = 0.8$ (80% of resting liquidity). Each pass determines an allocation as though resting orders could be arbitrarily overfilled, then the overfill quantity is reallocated among the remaining unfilled orders. The dot denotes the eventual limit. In this continuum approximation, the algorithm converges only in the limit of infinitely many steps, suggesting that in the real discrete case, convergence may require many iterations for large incoming market orders. The rate of convergence is slower for larger values of k .

with $q = Q/V$. If $qg(0) = k \leq 1$ then the allocation is determined in one pass. Figure 1 shows the result, for q small enough that $qg(0) = kq \leq 1$.

In Figure 1, the horizontal line $k = 1$ denotes the pro rata algorithm, in which each order receives the same fractional fill q . With the Euribor value $k = 2$, the earliest orders receive twice the allocation that they would with pure pro rata, and orders at the back of the queue receive very little. With the Short Sterling value $k = 4$, early orders receive four times as much as they would with pro rata, and orders at the back receive almost nothing. Since $g(1) = 0$ for any $k > 1$, these very small fills at the back of the queue are characteristic of the LIFFE fill algorithm.

Small values of q are by far the most reasonable for STIR products: Zhuo [2013] shows that mean trade sizes are typically smaller than one percent of quote size. Nonetheless, it is interesting to consider the full behavior of the match algorithm for large values of q .

Of course, if $q \geq 1$ then all resting orders are completely filled. If $1/k < q < 1$, then on the first pass orders are completely filled up to x_*^1 such that $qg(x_*^1) = 1$, and the algorithm is repeated with the remaining quantity, completely filling orders up to x_*^2, x_*^3, \dots and partially filling the remaining orders to varying degrees. In fact, in this continuous approximation, an *infinite* number of iterations will be required for convergence (Figure 2).

But since the functional form of the subsequent passes is exactly the same as



the initial function, the final overall allocation may be determined as

$$f(x) = \begin{cases} 1, & 0 \leq x \leq x_*(q) \\ \left(\frac{1-x}{1-x_*(q)}\right)^k, & x_*(q) \leq x \leq 1, \end{cases}$$

where $f(x)$ denotes the fraction of a limit order that is filled at position x in the queue, and $x_*(q) = \lim x_*^n$ is determined so that $\int_0^1 f(x) dx = q$.

The left column of Figure 3 illustrates the results of the LIFFE mixed match algorithm for large q . The top graph is the pro rata case $k = 1$, the bottom is the FIFO case $k = \infty$, and the intermediate graphs show the Euribor value $k = 2$ and the Short Sterling value $k = 4$. The different lines take $q = 0, 0.1, \dots, 0.9, 1$.

The highlighted lines illustrate the case $q = 0.6$. In the pro rata case $k = 1$, 60% of each order is filled, in the FIFO case $k = \infty$, the first 60% of resting volume is completely filled. With $k = 2$, the first 20% of resting volume is completely filled, and the remainder is partially filled with a linear profile. With $k = 4$, the first 46.7% of resting volume is completely filled, and the remainder is partially filled with a sharply decreasing density. Again, we see the characteristic behavior of the LIFFE algorithm: orders near the back receive extremely small allocations. In effect, the LIFFE algorithm is much more similar to a FIFO algorithm than to pro rata.

CME Mixed Match Algorithm

The CME interpolates between FIFO and pro rata algorithms by allocating the first p of the order using a FIFO method, and using pure pro rata for the remaining order size and the remaining unfilled orders. The 2-year Treasury futures contract uses 40% FIFO / 60% pro rata ($p = 0.4$), while the 10-year Treasury calendar spread contract uses a 20% FIFO / 80% pro rata mix ($p = 0.2$), beginning with the Feb. 2013 roll cycle. The CME has announced their intention to extend this mix to the other Treasury calendar spreads.

The right panel of Figure 1 shows this algorithm compared with the LIFFE algorithm; the parameters used in the picture are $p = 0.3$ and $q = 0.2$. The first $pq = 0.06$ of the resting orders are filled completely in the FIFO match. Then the remaining $(1-p)q$ of the incoming order is distributed uniformly over the remaining unfilled orders. If we neglect the overlap region in the lower left corner of size q^2 , then each order receives $(1-p)q$ partial fill.

The results for large q are shown in the right column of Figure 3, where the first figure is the percentage allocated to the FIFO match and the second figure is the percentage allocated to the pro rata match. The limit of a 0/100 mix (not shown), would match the pro rata picture in the top left, while the limit of a 100/0 mix (not shown) would match the FIFO picture in the bottom left.

Again, the example case $q = 0.6$ is highlighted. With a 20/80 mix, the first $0.2 * 0.6 = 0.12$ of resting volume is completely filled, and the remaining $0.8 * 0.6$ of incoming market order size is uniformly distributed across the remaining 0.88 of resting volume.



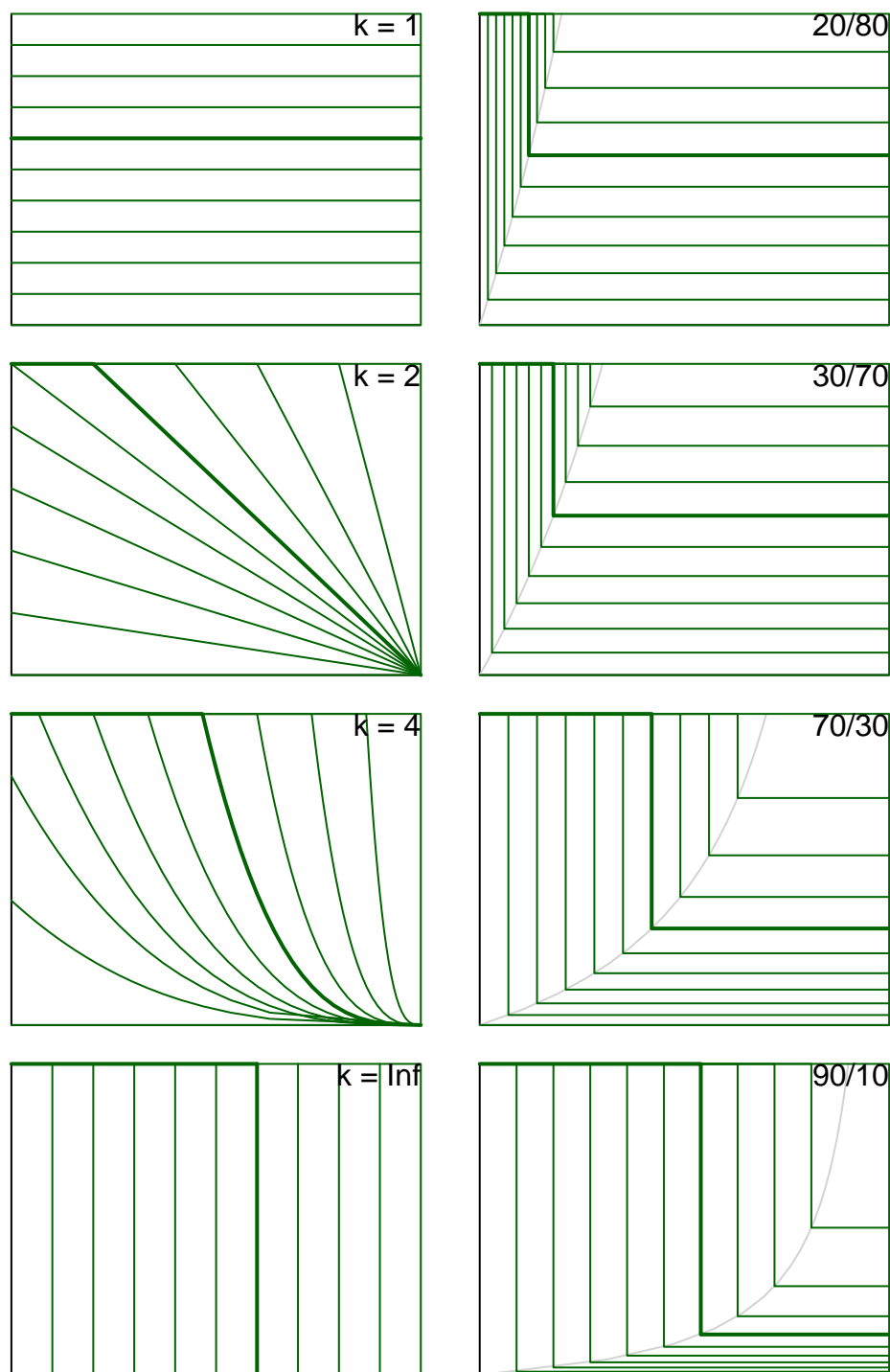


Figure 3: LIFFE weighted match algorithm (left column) and CME mixed algorithm (right). The curves represent incoming market orders 0%, 10%, 20%, ..., 90%, 100% of the total resting quote, with the example case 60% highlighted. As in Figure 1, horizontal axis is order position in the queue, and vertical axis is the fraction of an order that will be filled. Horizontal and vertical axis scales are 0 to 1.



Comparison of Match Algorithms

Both the LIFFE and the CME algorithms are an attempt to interpolate between the pure pro rata and the pure time priority (FIFO) algorithms. In our view, the LIFFE algorithm retains much more of the character of a FIFO algorithm, even for intermediate values of k . This is because the fill density $f(x)$ always drops to zero for orders at the back of the queue. Thus, even though late-arriving orders will not receive strictly zero quantity, the quantity received will be very small.

In contrast, the CME algorithm has nonzero density even at the right edge on the graphs in the right column of Figures 1 and 3. Thus even the last-arriving order will receive a nonzero percentage allocation. We believe that this is more effective at preserving both elements of FIFO and pro rata matching.

References

- J. Field and J. Large. Pro-rata matching and one-tick futures markets. Preprint, 2008.
- LIFFE. Changes to the Time Pro Rata Trade Matching Algorithm in Three Month Short Sterling and Three Month Euroswiss Interest Rate Futures Contracts. London Notice 3714, NYSE LIFFE, May 2013. Reversed May 30.
- J. McPartland. Recommendations for Equitable Allocation of Trades in High Frequency Trading Environments. Policy discussion paper, Federal Reserve Bank of Chicago, May 2013.
- Z. Zhuo. Characterization of futures by their market microstructure properties. Technical report, Quantitative Brokers, Feb. 2013.

Disclaimer

This document contains examples of hypothetical performance. No representation is being made that any account will or is likely to achieve profits or losses similar to those shown. There are frequently sharp differences between hypothetical performance results and the actual results subsequently achieved by any particular trading program.

Hypothetical performance results are prepared with the benefit of hindsight. Hypothetical trading does not involve financial risk, and no hypothetical trading record can completely account for the impact of financial risk in actual trading. The ability to withstand losses or to adhere to a particular trading program in spite of trading losses are material points which can also adversely affect actual trading results. Numerous other factors related to the markets in general or to the implementation of any specific trading program cannot be fully accounted for in the preparation of hypothetical performance results and can adversely affect actual trading results.

Futures are speculative products and the risk of loss can be substantial. Futures spreads are not necessarily less risky than short or long futures positions. Consequently, only risk capital should be used to trade futures. The information contained herein is based on sources that we believe to be reliable, but we do not represent that it is accurate or complete. Nothing contained herein should be considered as an offer to sell or a solicitation of an offer to buy any financial instruments discussed herein. All references to prices and yields are subject to change without notice. Past performance/profits are not necessarily indicative of future results. Any opinions expressed herein are solely those of the author. As such, they may differ in material respects from those of, or expressed or published by or on behalf of, Quantitative Brokers or its officers, directors, employees or affiliates.

