The Need for (Trading) Speed

BY JESSIE ROMERO

High-frequency trading (HFT) — using powerful computers and complex algorithms to trade securities at very fast speeds — is the subject of heated debate. Defenders of HFT say it benefits investors by making markets more efficient and more liquid. Critics worry it makes financial markets unstable and stacks the deck against investors who can't afford to invest in high-speed infrastructure.

Those investments are considerable. In 2010, a privately built $300 million high-speed fiber-optic cable reduced the transmission time between Chicago and New York from 16 milliseconds to 13 milliseconds; according to some reports, trading firms paid as much as $300,000 per month for access. (A millisecond is one one-thousandth of a second.) Just a few years later, fiber-optic cable seems obsolete, as trading firms have begun using microwave towers and laser beams to shave off additional milliseconds. Other firms pay high fees to locate their servers in the same facilities as securities exchanges' servers; the exchanges measure carefully to make sure one firm's cord isn't a few feet shorter than another's.

In a recent article, Eric Budish and John Shim of the University of Chicago and Peter Cramton of the University of Maryland conclude that this "arms race" for speed is the inevitable result of the market design, which treats time as continuous rather than discrete. In general, exchanges operate based on a limit order book, which constantly matches "bids" to buy a security with "asks" to sell a security. Orders are accepted based on price-time priority: Bids or asks with the most attractive price are accepted first, and ties are broken based on when the order was received. But treating time as continuous creates mechanical opportunities for arbitrage, according to the authors, and gives firms an incentive to invest in speed.

To demonstrate this, Budish, Cramton, and Shim begin by examining data on two securities, the E-mini S&P 500 Futures Contract (ticker ES) and the SPDR S&P 500 exchange traded fund (ticker SPY), between Jan. 1, 2005, and Dec. 31, 2011. The authors find they are nearly perfectly correlated over relatively long intervals, such as a minute, hour, or day. But at higher-frequency intervals, such as a millisecond, the correlation breaks down completely. This creates opportunities for arbitrage: If a high-frequency trading algorithm observes an increase in the price of ES, for example, it can sell ES and buy SPY before the price of SPY has time to change. And since someone is always first, this creates an incentive to be the fastest.

In theory, arbitrage opportunities don't last; once the market discovers them, competition will cause prices to converge. But Budish, Cramton, and Shim find that while the duration of arbitrage opportunities shrank significantly over the course of their study, from a median of 97 milliseconds in 2005 to a median of 7 milliseconds in 2011, the profitability of arbitrage opportunities stayed constant. They write, "The arms race does not actually affect the size of the arbitrage prize; rather, it just continually raises the bar for how fast one has to be to capture a piece of the prize."

To account for their empirical findings, the authors construct a simple model in which investors and trading firms buy and sell a security and receive public signals about that security's value (such as the latest price of a correlated security). When there is a change in the signal, a trading firm sends a message to the exchange asking it to cancel its existing quotes for the security and to replace them with new quotes. At the exact same time, however, other trading firms try to "snipe" the stale quote; that is, they send a message to the exchange to buy or sell the security at the old price. Because the exchange processes the orders serially, there is a high probability that the initial trading firm gets sniped even though all the firms observed the signal at the same time. This raises the cost of providing liquidity to investors, since trading firms must charge a higher bid-ask spread to cover the risk of being sniped.

When the authors modify their model to allow trading firms to invest in speed, the equilibrium result is a socially wasteful arms race. Some firms invest in speed to be the first to snipe, other firms invest to avoid being sniped, and because competition does not eliminate the arbitrage opportunities, the incentive is to continue investing. At the same time, competition dissipates the net rents trading firms can earn, and investors ultimately bear the cost of speed in the form of higher liquidity costs.

Budish, Cramton, and Shim propose ending the arms race by holding batch auctions at discrete intervals, such as 100 milliseconds, rather than processing orders serially. In their model, batch auctions significantly reduce the value of slight speed advantages and eliminate the rents trading firms can earn on symmetrically observed public information. Unlike other proposals to curb the HFT arms race, such as taxes, minimum resting times for quotes, or random delays in processing messages to the exchanges, batch auctions that treat time as discrete rather than continuous could address what the authors view as the fundamental flaw in the system. **EF**

"The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." By Eric Budish, Peter Cramton, and John Shim. *Quarterly Journal of Economics*, November 2015, vol. 130, no. 4, pp. 1547-1621.