

April 22, 2016

David R. Pearl
Office of the Executive Secretary
U.S. Department of the Treasury
1500 Pennsylvania Avenue, NW
Washington, D.C. 20220

Re: Notice Seeking Public Comment on the Evolution of the Treasury Market Structure (Docket No. TREAS-DO-2015-0013)

Dear Mr. Pearl:

I appreciate the opportunity to comment on the topic of the Evolution of the Treasury Market Structure. I am an economics professor at the University of Chicago Booth School of Business who researches market design – designing the “rules of the game” for markets – with a specific focus on the design of financial exchanges. Market design research assumes that participants in a market act optimally in their rational self-interest with respect to market rules, but takes seriously the possibility that the market rules themselves may be sub-optimal. I believe that this approach brings a useful perspective to the discussion about regulation of the secondary market for Treasuries. I write independently, without any financial involvement with any of the participants in this discussion.

The RFI covers a broad range of important topics. I will focus the bulk of my letter on one structural issue with the design of modern electronic financial exchanges. This structural issue – essentially, a mathematical glitch introduced in the transition from human-based markets to electronic markets – relates to many of the themes of the RFI: liquidity, transparency, computational complexity, timestamp accuracy, and the role of high-frequency trading.¹

The structural issue is that modern electronic financial exchanges treat time as *continuous* – meaning, for any two orders received, one is first, even if by the tiniest imaginable fraction of a second (e.g., 0.000000001 seconds) – and process messages *serially*, that is, one-at-a-time in order of receipt. Treating time as continuous, and then processing messages serially upon receipt, causes harmful latency arbitrage to be an intrinsic “built in” feature of the market design, which both harms liquidity and induces a never-ending race for speed. Treating time as continuous also complicates the market computationally – modern computers and communications links are fast but not infinitely fast, so computers and communications are intrinsically latent in a continuous market. Last, treating time as continuous complicates the market paper trail and market surveillance, both because timestamps cannot be perfectly accurate and because relativity becomes relevant if time is measured finely enough.

¹ The specific items in the RFI my remarks relate most closely to are 1.1 (liquidity), 1.5 (market structure), 1.6 (speed), 2.1 (risk management), 2.3 (algorithmic trading), 3.1 (monitoring), and 3.7 (timestamp accuracy).

The alternative approach suggested by my research is to treat time as *discrete*. Discrete time means that time is put into units, much the same way that prices come in units like \$0.01 or 1/64th. The time interval should be long compared to the time it takes information to travel between exchanges, and long compared to the time it takes exchange computers to perform simple processing and communications tasks, but otherwise can be extremely short. For instance, 0.01 seconds would be long relative to latencies between futures markets in Chicago and cash markets in New York / New Jersey, and would also be long relative to the time it takes exchange computers to perform simple tasks.

Once time is put into discrete units, it becomes possible that multiple orders arrive to the exchange at the same (discrete) time; e.g., within the same 0.01 seconds. In such a case, my research suggests that such orders arriving at the same discrete time be processed in *batch* (as opposed to serially), using a uniform-price auction. This market design, in which time is discrete and orders arriving at the same time are processed in batch, is called *frequent batch auctions*. In a typical short time interval, when there is little or no activity, the market design behaves almost identically to the electronic limit order book. But, in time intervals when there is a burst of activity it is importantly different because of the use of an auction – in such instances it is more analogous to the market design the Treasury currently uses for the primary market for Treasury securities.

Frequent batch auctions have five important benefits for markets relative to continuous trading. First, they eliminate latency arbitrage, by transforming competition on speed into competition on price. Intuitively, if there is new information that many algorithmic trading firms observe at roughly the same time, firms will compete over who is willing to pay the most or sell for the least instead of competing over who can respond the fastest to snipe stale quotes. Second, by eliminating latency arbitrage, they enhance liquidity. Intuitively, it is easier to provide liquidity if you are not worried about being 0.00000001 seconds late to change your prices in response to some news event to avoid being sniped. Third, frequent batch auctions stop the never-ending race for speed. Intuitively, if trading occurs every 0.001 or 0.01 seconds, then saving 0.000001 seconds (considered a meaningful speed improvement these days) is much less valuable. Fourth, frequent batch auctions simplify the market computationally. Discrete time respects the limits of computers and communications, whereas continuous time trading implicitly assumes that computers and communications are infinitely fast. Fifth, frequent batch auctions simplify the market's paper trail. Issues like clock synchronization and relativity that complicate the market paper trail in continuous time become irrelevant in discrete time.²

I emphasize that while my research is critical of some aspects of high-frequency trading, my research does not suggest that we should be nostalgic for the days of human-based trading. The empirical record is clear that the transition from human-based markets to electronic markets has on the whole been quite positive, both for investors and for overall market efficiency, and that algorithmic trading plays a

² For the full details of the argument against continuous-time trading and in support of discrete-time trading, see Eric Budish, Peter Cramton and John Shim, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *Quarterly Journal of Economics*, Vol. 130(4), November 2015, pgs. 1547-1621. Available under Open Access license at: <http://faculty.chicagobooth.edu/eric.budish/research/HFT-FrequentBatchAuctions.pdf>.

useful role in price discovery and liquidity provision.³ Rather, my research suggests that the current design of electronic financial exchanges leaves meaningful room for improvement.

I encourage the Treasury to conduct a pilot test of frequent batch auctions. There are a number of important details that would be critical to get right in such a pilot test. One detail that I will underscore here is that if conducted in the context of a market that is otherwise continuous I would suggest an extremely fast batch interval, such as 1 millisecond. A small pilot test would not make a dent on the overall speed race; rather, the goal would be to eliminate sniping and the harm to liquidity provision and to measure these effects.

I also encourage the Treasury to ensure that there are not regulatory obstacles to private adoption of frequent batch auctions by an exchange, whether an existing exchange or a new venture. In equities markets, Regulation National Market System has constrained market design innovation, though there are some recent signals that the SEC is looking to relax those constraints. I urge the Treasury to ensure that it does not similarly constrain market design innovation.

Last, I encourage the Treasury to make precisely-timestamped financial market data more widely available for research. According to McKay Brothers co-founder Stephane Tyc, writing in Sept 2014, “it is now routinely possible to synchronize [exchange matching engine clocks] to about one microsecond.”⁴ Tyc goes on to suggest that a mandate that clocks be synchronized to within about 10 microseconds would be reasonable, and I agree with the spirit of this: if exchanges operate in continuous time it is impossible to provide perfectly accurate timestamps but they should at least provide timestamps that are as accurate as technologically possible. Ideally, precisely-timestamped data should be made available both to investors so that they can assess execution quality (e.g., did they repeatedly trade at a price that 10 or 100 microseconds later would have been more advantageous), and to academics for research. For example, to fully study the extent of latency arbitrage in Treasury markets would require precisely-timestamped data from all of the major components of the secondary market for Treasuries: the major cash trading platforms, the dealer-to-customer cash market, and the futures market.

I will be pleased to be of service to the Treasury in these important matters in whatever way is helpful.

Kind regards,

Eric Budish

³ For overall time series evidence, see Andrea Frazzini, Ronen Israel and Tobias J. Moskowitz, “Trading Costs of Asset Pricing Anomalies,” Chicago Booth Research Paper No. 14-05, September 2015. For a study of the transition from human-based trading to algorithmic trading, see Terrence Hendershott, Charles Jones, and Albert Menkveld, “Does Algorithmic Trading Improve Liquidity?” *Journal of Finance*, Vol. 66(1), February 2011, pgs. 1-33. For a survey of related research, see Charles Jones, “What Do We Know About High-Frequency Trading?”, Columbia Business School Working Paper, March 2013.

⁴ Stephane Tyc, 2014, “A Technological Solution to Best Execution and Excessive Market Complexity.” Available at <http://www.quincy-data.com/wp-content/uploads/2014/12/20141002-Atechsolutiontobestexec.pdf?aad152>