A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?

[ONLINE APPENDICES]

Journal of Political Economy

Eric Budish, Robin S. Lee, and John J. Shim

A   Supporting Details on Exchange Trading Fees

Exchange trading fees are notoriously complicated. For example, Figure A.1, which comes from an investment bank research report, provides a tongue-in-cheek depiction of the set of trading fees faced by a trader depending on a variety of factors (exchange, order type, participant type, etc.). Our task is to cut through this complexity to report per-share per-side regular hours trading fees that are representative of what participants typically pay. We do this two ways: by estimating average fees (Appendix A.1) and collecting historical fee schedules (Appendix A.2).

A.1   Details for Average Trading Fee Calculations

In this appendix, we provide supporting details for the calculation of average per-share per-side trading fees for the three major exchange families as reported in Table 2.1. The calculations themselves are available in a supporting spreadsheet available in the online appendix.

BATS.   For BATS, the April 2016 S-1 provides a net trading revenue figure of $81.0M and a matched share volume figure of 1.5 billion shares per day which corresponds to 378 billion shares per year (252 trading days). We cross-checked the volume figure with the NYSE TAQ data and found 367.9 billion shares in that data set, which is within rounding error. Using the S-1 figures for consistency with what follows, we obtain net revenue per share of $81M / 378 billion shares = $0.000214 which corresponds to $0.000107 per-share per side. This figure includes revenue from regular-hours trading, which is what we want, but it also includes revenue from opening and closing auctions and routing, which we want to strip out.

For BATS, the auction volume is minimal (0.13 billion shares per NYSE TAQ), so even under the assumption that all auction volume pays the maximum auction fee (which, depending on the order type utilized, ranges from zero to $0.0005, or 5 mills, for the opening auction and $0.0010, or 10 mills, for the closing auction), auction revenue does not move the needle. Routing volume on the other hand is significant, at approximately 25.2 billion shares per the S-1 (0.1 billion per day times 252 trading days). BATS reports routing and clearing

Figure A.1: Complexity of Exchange Fees
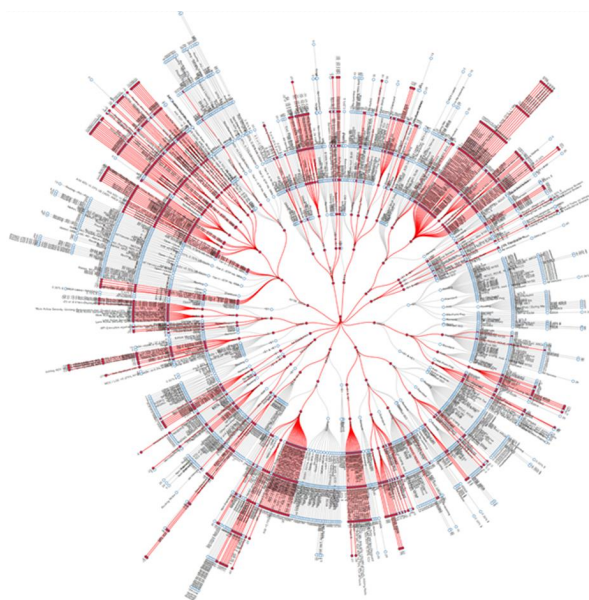
Notes: From RBC Capital Markets, "Complexity of Exchange Pricing and Corresponding Challenges to Transparency and Routing," (February 2016). RBC's visual representation of exchange trading and routing fees by tier as of October 2015. Each path from the center node to an extreme node indicates a particular fee scenario, characterized by exchange, share price, transaction type (adding or removing liquidity), market participant type, market participant ADV, security listing exchange, other factors, and a fee or rebate. Red indicates rebate, blue indicates fee.

costs of $43.7M in their S-1, which is 17.3 mills per share. We use a variety of data regarding routing fees based on the ultimate destination of the trade (e.g., a directed ISO versus a take on another exchange versus a take on a dark venue) to obtain a back-of-envelope estimate for BATS's routing revenue per share of 22.8 mills per share and hence net routing revenue of 5.5 mills per share. This in turn implies net routing revenue for BATS overall of 5.5 mills * 25.2 billion shares = $13.8 million. Subtracting this revenue from BATS's total revenue as reported above yields $67.2 million of regular-hours trading revenue, or $0.000178 per share and $0.000089 per-share per-side, as reported in Table 2.1. We caveat that the routing estimate is particularly back-of-envelope, so the reader may prefer to utilize the $0.000107 figure reported above or to adjust for routing in some other way.

Nasdaq. Nasdaq last reported U.S. cash equity net trading revenues in 2013; in 2015, they only report equity net trading revenues globally, not for the U.S. Our first step therefore is to take the 2015 number for global cash equity trading revenues less transaction-based expenses (i.e., rebates), of $253 million, and multiply it by the 2013 ratio of U.S.:Global equity trading revenues, which was $107M/$193M = 55%. This yields $140.3M for 2015 U.S. cash equity net trading revenues. Nasdaq reports matched U.S. equity share volume of 327.7 billion; this is close to the figure we obtain in the TAQ as a cross-check (329.4 billion shares). We thus obtain net revenue per share of $140.3M / 327.7 billion shares = $0.000428, or $0.000214 per-share per-side. We caveat that this figure will be incorrect if the 2015 U.S.-to-Global ratio is meaningfully different from the 2013 ratio.

The next step is to deduct auction volume and revenue, which are both significant. We obtain auction

volume from TAQ, of 5.3 billion shares annually for the opening auction and 20.2 billion shares for the closing auction. For the opening auction, we use a fee of 15 mills per-share per-side, which is the fee for regular market-on-open and limit-on-open orders that participate in the auction. This ignores fees for some other less-common order types as well as a fee cap for high-volume users of $20,000 per month. For the closing auction, Nasdaq has a fee schedule with 6 tiers based on volume levels. The fee ranges from 8 mills for the highest-volume tier to 15 mills for the lowest. We assume an equal six-way split across the six tiers to obtain 12.1 mills. Together, the opening and closing auction account for 25.5 billion shares traded and $64.6 million of revenue.

Last, we deduct routing revenue. Routing is prominently discussed in Nasdaq financial statements but they do not report any specific numbers. Since the Nasdaq routing business appears to be at least somewhat similar to the BATS routing business, we utilize the BATS net routing revenue per share number computed above (5.5 mills) and the BATS routed volume as a % of total volume (6.7%), to obtain net routing revenue of $12.0 million on 21.8 billion shares.

When we subtract auction revenue and volume, and subtract routing revenue, we obtain 302.2 billion regular-hours shares traded and $63.6 million of regular-hours net trading revenue, for $0.000211 per share and $0.000105 per-share per-side, as reported in Table 2.1. As a sensitivity analysis, we assume that we have overestimated auction revenues by 25%, for example, due to the monthly fee caps. This would change the figure to $0.000132 per-share per-side.

NYSE NYSE's parent company, Intercontinental Exchange (ICE), reports in its 2015 10-K that NYSE's U.S. cash equities revenues, net of transaction based expenses (i.e., rebates), were $220 million in 2015. The ICE 10-K reports average daily matched volume of 1,187M shares for Tape A, 296M shares for Tape B, and 206M for Tape C. Multiplied by 252 trading days this yields annual volume of 425.6 billion shares, which is close to the TAQ number. This yields revenue per share of $220M / 425.6 billion shares = $0.000517, or $0.000258 per-share per-side.

Next, we deduct auction revenue and volume. We get opening and closing auction volume for NYSE, NYSE Arca, and NYSE Mkt from the TAQ data. These volumes are significant for both NYSE and NYSE Arca, with 11.1 billion and 1.9 billion shares of volume for the open, and 48.4 billion and 9.7 billion shares of volume for the close, respectively. For the opening auction, we use a fee of 10 mills for NYSE and NYSE Mkt and 15 mills for NYSE Arca, based on their fee schedules. As with Nasdaq, there are some discounts (in particular for NYSE designated market makers) and monthly caps, which we do not attempt to account for here, but rather do so below in a sensitivity analysis. For the closing auction, NYSE has a range of fees from 6 mills to 10 mills depending on volume tier; we use an equal-weighted average of the tiers to obtain 7.7 mills. NYSE Arca's closing auction fee is 10 mills and NYSE Mkt's is 8.5 mills. Combined across these three venues and combining both the open and close, we obtain $123.3M of total auction revenue.

For routed volume, we utilize that the ICE 10-K reports both matched volume and handled volume; the difference is what is routed. This comes to 10.8 billion shares annualized across the 3 tapes. We utilize the same 5.5 mills net routing fee number from BATS, lacking any better source. This comes to $5.9M of total routing revenue.

When we subtract auction revenue and volume, and subtract routing revenue, we obtain 353.5 billion regular-hours shares traded and $90.7 million of regular-hours net trading revenue, for $0.000257 per share and $0.000128 per-share per-side, as reported in Table 2.1. As a sensitivity analysis, we assume that we have overestimated auction revenues by 25%, for example, due to the monthly fee caps. This would change the figure to $0.000172 per-share per-side.

Table A.1: U.S. Equity Exchange Trading Fees ("f")

Exchange	Fee Type	Taker Fee		Maker Fee		Total fee per share per side	
		Min	Max	Min	Max	Min	Max
NASDAQ	Maker-Taker	0.00300	0.00300	-0.00305	-0.00150	-0.00003	0.00075
BATS BZX	Maker-Taker	0.00300	0.00300	-0.00320	-0.00200	-0.00010	0.00050
EDGX	Maker-Taker	0.00300	0.00300	-0.00350	-0.00200	-0.00025	0.00050
NYSE	Maker-Taker	0.00270	0.00270	-0.00220	-0.00140	0.00025	0.00065
NYSE Arca	Maker-Taker	0.00280	0.00300	-0.00300	-0.00200	-0.00010	0.00050
BATS BYX	Taker-Maker	-0.00160	-0.00160	0.00140	0.00180	-0.00010	0.00010
EDGA	Taker-Maker	-0.00020	-0.00020	0.00030	0.00050	0.00005	0.00015
NASDAQ BX	Taker-Maker	-0.00150	-0.00040	0.00165	0.00200	0.00008	0.00080

Notes: The table summarizes the fee schedules for the top 8 exchanges retrieved from the Internet Archive (Wayback Machine) dated from February 28, 2015 to September 1, 2015 . In general, we determine the max rebates based on what a trading firm that satisfies the exchange's highest volume tier would pay or receive, and the min rebates and fees tend to be the baseline for adding or taking liquidity. We omit fees associated with special programs or differences based on tape plans. Please see Appendix A.2 for a complete table of estimated fees for both regular and special programs and for tape A, B, and C stocks.

A.2 Range of U.S. Equity Exchange Trading Fees

In this appendix, we use historical fee schedules from exchange websites retrieved using the Internet Archive to report the range of trading fees for the top 8 exchanges. We focus on 2015 for consistency with the other analyses.[1]

Table A.1 presents the observed range for per-share per-side regular-hours trading fees for the top 8 exchanges. As can be seen, many of the exchanges have minimum fees on a per-share per-side basis that are actually slightly negative. Table A.2 provides a more complete table, which takes into account additional details for special fee programs as well as the fee ranges by listing exchange (referred to as the "Tape" a stock is included on: Tape "A" represents stocks listed on NYSE, "B" on NYSE Arca, and "C" on Nasdaq). The more complete table shows that 7 of the 8 exchanges have a minimum per-share per-side fee that is negative. The only exception is EDGA, which has a minimum total fee per share per side of $0.00005 of 0.5 mills. The maximum fee per-side is always strictly positive and typically about $0.0005 or 5 mills. Roughly, low-volume market participants pay the maximum fee and the highest-volume market participants pay the minimum fee.

For the purpose of considering whether fees are negative enough to generate profits from trading with oneself, we should look at the SEC + FINRA fees on a per-share per-side basis because an exploiter of a self-trading strategy would need to both buy and sell. For a $5 stock this would be $0.000114 or 1.14 mills, i.e., per-share per-side fees could go to -1.14 mills without making self-trading profitable.[2] This may help explain why exchange trading fees are able to go slightly negative without making self-trading profitable.

---

[1]The specific historical fee schedules we use range from Feb 2015 to Sept 2015 depending on the Internet Archive's coverage. The specific financial filings we use are the BATS April 2016 S-1 filing, Nasdaq's fiscal year 2015 10-K report, Intercontinental Exchange's (NYSE's parent) fiscal year 2015 10-K report, and NYSE's fiscal year 2012 10-K filing (2012 was its last full fiscal year as a stand-alone company).

[2]At the time of our data, the SEC fee was $21.80 per $1M traded and the FINRA fee was $0.000119 per share traded. Both fees are assessed on sales but not purchases, i.e., they are assessed on one side of each transaction. For the purpose of calculating the self-trading boundary, we look at the SEC + FINRA fees on a per-share per-side basis because an

Table A.2: U.S. Equity Exchange Trading Fees Per Share ("f")

Exchange	Fee Type	Program	Tape	Maker Fee		Taker Fee		Total fee per share		Total fee per share per side	
				Min	Max	Min	Max	Min	Max	Min	Max
NASDAQ	Maker-Taker	Regular	C	-0.00305	-0.00150	0.00300	0.00300	-0.00005	0.00150	-0.00003	0.00075
NASDAQ	Maker-Taker	DLP	C	-0.00400		0.00300	0.00300	-0.00100		-0.00050	
NASDAQ	Maker-Taker	Regular	A/B	-0.00305	-0.00200	0.00295	0.00300	-0.00010	0.00100	-0.00005	0.00050
BATS BZX	Maker-Taker	Regular	NA	-0.00320	-0.00200	0.00300	0.00300	-0.00020	0.00100	-0.00010	0.00050
BATS BZX	Maker-Taker	NBBO Setter	NA	-0.00360		0.00300	0.00300	-0.00060		-0.00030	
EDGX	Maker-Taker	Regular	NA	-0.00350	-0.00200	0.00300	0.00300	-0.00050	0.00100	-0.00025	0.00050
NYSE	Maker-Taker	Regular	A	-0.00220	-0.00140	0.00270	0.00270	0.00050	0.00130	0.00025	0.00065
NYSE	Maker-Taker	SLP	A	-0.00290		0.00270	0.00270	-0.00020		-0.00010	
NYSE	Maker-Taker	DMM	A	-0.00340	-0.00150	0.00270	0.00270	-0.00070	0.00120	-0.00035	0.00060
NYSE Arca	Maker-Taker	Regular	B	-0.00230	-0.00200	0.00280	0.00300	0.00050	0.00100	0.00025	0.00050
NYSE Arca	Maker-Taker	LMM	B	-0.00450		0.00250	0.00250	-0.00200		-0.00100	
NYSE Arca	Maker-Taker	Regular	A/C	-0.00300	-0.00200	0.00300	0.00300	0.00000	0.00100	0.00000	0.00050
BATS BYX	Taker-Maker	Regular	NA	0.00140	0.00180	-0.00160	-0.00160	-0.00020	0.00020	-0.00010	0.00010
BATS BYX	Taker-Maker	NBBO Setter	NA	0.00130		-0.00160	-0.00160	-0.00030		-0.00015	
EDGA	Taker-Maker	Regular	NA	0.00030	0.00050	-0.00020	-0.00020	0.00010	0.00030	0.00005	0.00015
NASDAQ BX	Taker-Maker	Regular	NA	0.00165	0.00200	-0.00150	-0.00040	0.00015	0.00160	0.00008	0.00080
NASDAQ BX	Taker-Maker	QMM	NA	0.00140		-0.00150	-0.00040	-0.00010		-0.00005	

Notes: This table summarizes the fee schedules for the top 8 exchanges retrieved from Internet Archive (Wayback Machine) dated from February 28, 2015 to September 1, 2015 (BATS Global Markets, Inc., 2015*a,b,c,d*; Nasdaq, Inc., 2015*a,b*; NYSE, 2015*a,b*; NYSE Arca Equities, Inc., 2015). In general, we determine the max rebates based on what a trading firm that satisfies the exchange's highest volume tier would pay or receive, and the min rebates and fees tend to be the baseline for adding or taking liquidity. We consider all volume-based incentives for regular-hours liquidity provision, but we do not include additional incentives for trading off-hours, trading at the open or close, creating non-displayed midpoint liquidity, sending retail orders, routing, or for trading securities with a share price below $1. The "Regular" program corresponds to the fees and rebates a firm would receive if it does not qualify for additional incentive programs detailed below, which often either involve an additional volume threshold, a National Best Bid and Offer quoting requirement, or an off-hours trading requirement. The Designated Liquidity Provider (Nasdaq DLP) program rewards market participants who maintain a one or two-sided quote on specified Nasdaq-listed ETFs for at least 15% of the trading day. The National Best Bid or Offer Setter (BZX/BYX NBBO Setter) program rewards participants who send orders that set the new national best bid or offer, as well as fulfill an additional volume requirement. The Supplemental Liquidity Provider (NYSE SLP) program rewards participants who quote at the NBBO at least 10% of the trading day, as well as fulfill an additional volume requirement. The Designated Market Maker (NYSE DMM) program rewards participants who make commitments to satisfy a wide variety of requirements involving market depth, volume, NBBO quoting, capital, and others every month. The Qualified Market Maker (Nasdaq BX QMM) program grants a discount on making liquidity for participants who actively quote at the NBBO. We also separately report fees by "Tape" or listing exchange. Tape "A" represents stocks listed on NYSE, "B" on NYSE Arca, and "C" on Nasdaq. In the table, NA indicates that an exchange does not charge different fees by listing exchange.

B   Supporting Details for Exchange Speed-Technology Revenue

B.1   Details for 2015 Data and Co-Location Revenue Estimates for Nasdaq and NYSE

In this appendix, we provide supporting details for our calculations of market data and co-location/connectivity revenues for Nasdaq and NYSE in 2015, which we reported in Section 2.2. BATS's 2015 revenue figures come directly from filings. Specifically, BATS's April 2016 IPO filing (i.e., form S-1) provides an unusually clear window into how exchange revenues break down across trading revenue, market data, and co-location/connectivity. BATS was acquired by CBOE later in 2016 and following that acquisition no longer reported their revenues with such granularity. Thus, we focus on 2015 revenues for this analysis (note: neither Nasdaq nor NYSE have ever reported their U.S. equities revenue with the granularity of BATS's IPO filing).

Nasdaq's fiscal year 2015 10-K reports market data and co-location/connectivity revenue only at the global level—$399M and $239M, respectively.[3] To get from global to the U.S., for market data, we utilize information in its 2013 10-K filing that breaks out its market data business geographically: U.S. is 72% of the total in 2013, and we assume this ratio holds in 2015. For co-location/connectivity, we use Nasdaq's overall 2015 U.S.:Global revenue ratio, of 71%. Last, we need to separate out Nasdaq's U.S. Equities business from its U.S. Options business. We take two approaches. First, we assume that Nasdaq's market data and co-location revenue from U.S. Equities vs. U.S. Options is proportional to its trading volume in U.S. Equities vs. U.S. Options. Second, we assume that Nasdaq's U.S. Options business generates the same market data and co-location revenue as BATS's U.S. Options business, scaled up for Nasdaq's larger U.S. Options volume than BATS. The first approach assumes that every 1 option traded on Nasdaq generates the same market data and co-location revenue as 100 shares of stock; the second approach assumes that 1 option traded on Nasdaq generates the same market data and co-location/connectivity revenue as 1 option traded on BATS. These two approaches yield a range for Nasdaq's U.S. Equities revenue of $222.4M-$267.3M for market data, $121.0M-$139.0M for co-location/connectivity, and $343.3M-$406.4M combined.

NYSE was acquired by ICE, a large futures exchange conglomerate, in Nov 2013. ICE's 2014 10-K filing therefore gives significant detail on the contribution of the NYSE business to the overall ICE business, for 2014, the first full year of integration (and also for the Nov-Dec 2013 period). The filing reports that NYSE's U.S. businesses (not including Euronext, which ICE divested) contributed $430M to its data services business in 2014; this includes both market data and co-location/connectivity, for both U.S. equities and U.S. options. The filing also reports that $202M of this was for co-location/connectivity, implying $228M for market data. ICE's 2015 10-K filing reports that it reclassified an additional $60M of revenue, for 2014, from its "other" category to its data services business, and that this revenue corresponds to "NYSE connectivity fees and colocation service revenues".[4] Therefore the adjusted 2014 totals are $262M for co-location/connectivity and $228M for market data. Comparison of ICE's 2014 and 2015 10-K filings suggest a growth rate of its overall data services business from 2014 to 2015, of which the NYSE business was by far the largest component, of 12.3%.[5] For comparison,

---

exploiter of a self-trading strategy would need to both buy and sell.

[3]The $239M for global co-location/connectivity also contains revenues from a small Nordic region Broker Services business, which when last reported separately was $19M; we subtract out this $19M from Nasdaq's global "Access and Broker Services" business in the analysis that follows.

[4]It is hard to know but we guess that this adjustment reflects post-merger alignment of accounting practices between NYSE and ICE, that in principle should have been reflected in the 2014 10-K but that was not completed until the 2015 10-K.

[5]This growth figure accounts for several other ICE acquisitions in this time period. 2014 ICE data services revenue

BATS's U.S. equities growth rate for the 2014 to 2015 period was 19.2% for co-location/connectivity and 12.4% for market data,[6] which suggests that the 12.3% growth rate computed from ICE data is reasonable for NYSE. We use this growth rate to obtain estimates for 2015 for NYSE's overall U.S. business, and then utilize the same two methods described above for Nasdaq to obtain estimates for NYSE's U.S. Equities business. This yields a range of \$218.9-\$241.5M for U.S. equities market data, \$251.6-\$281.5M for U.S. equities co-lo/connectivity, and \$470.5-\$523.0M combined.

B.2 Details for Data and Co-Location Revenue Growth Estimates for Nasdaq and BATS

In this appendix we provide supporting details for the data on Nasdaq and BATS exchange-specific speed technology (ESST) revenue growth. Our goal is to get a sense of magnitudes for ESST growth over time by looking at revenue growth in the financial reporting categories that contain U.S. equities ESST revenues. NYSE's financial reporting segments changed too frequently in the Reg NMS era for the exercise to be instructive. This is likely due to NYSE's merger with Euronext in 2007, its acquisition by ICE in 2013, and some financial segment reporting changes in the period in between. With this caveat in mind, the category "Technology Services Revenues" which includes co-location/connectivity grew from \$137M to \$341M over the period 2006-2012, and the category "Market Data Revenues" grew from \$223M to \$348M over this time period.

From 2006 to 2012, Nasdaq reported co-location/connectivity revenues in the category "Access services revenues." In 2013, Nasdaq changed the reporting category to "Access and Broker Services Revenues," which incorporated Nasdaq's small Nordic broker services business (\$19M in 2012) into the category. In 2015, Nasdaq appears to have adjusted its accounting practices to reclassify some revenue in "Access and Broker Services Revenues" to "Technology Solutions," which led to a downward revision of \$18M for the 2014 revenue figures based on the 2015 reporting method as reported in Nasdaq's fiscal year 2015 10-K, so it is of a similar magnitude as the upward revision in 2013. In 2016, Nasdaq changed the reporting category again to "Trade Management Services Revenues," but this appears to be a change in the category name only with no revision to revenue figures reported in previous 10-Ks. We view the periods 2006-2012 and 2015-2017 as yielding reliable apples-to-apples growth rates, and the period 2012-2015 seems relatively flat with the caveat that there were multiple reporting changes. We find the Nasdaq reporting categories containing co-location and connectivity revenue quadrupled in the 2006-12 period (growth of 26.7% per year), was roughly flat in the 2012-2015 period, and then in 2015-2017 growth was 10.3% per year.

From 2006 to 2012, Nasdaq reported its U.S. equities proprietary market data revenue in the category "U.S. market data products." Over this period, Nasdaq also separately reports tape revenues as "Net U.S. tape plans." Starting in 2013, Nasdaq reports only combined U.S. data revenue (i.e., including both proprietary and tape) in the segment "U.S. market data products," and also made some segment reporting changes, moving \$27M of revenue from "Index data products" out of U.S. market data products into its own reporting category. Starting in 2014, Nasdaq reports only total market data revenue instead of separating out U.S., international, and index

---

was \$691M but includes just 12 weeks of the SuperDerivatives business, which contributed \$12M in those 12 weeks; therefore 2014 revenue pro forma for the SuperDerivatives business was \$731M. 2015 data services revenue was \$871M but includes \$50M of revenues from 2015 acquisitions of Interactive Data and Trayport; therefore a like-for-like 2015 revenue number is \$821M, or 12.3% more than the adjusted 2014 figure.

[6]BATS's 2014 numbers include just 11 months of Direct Edge revenue versus 12 months in 2015. If we conservatively assume that the Direct Edge business is 50% of BATS's overall business, then we can take the unadjusted 2014-to-2015 growth rates, of 23.7% for co-location/connectivity and 16.9% for market data, and reduce them by $50\% \cdot \frac{1}{11} \approx 4.5$ percentage points, to obtain 2014 to 2015 growth rates that are apples-to-apples.

market data separately. To get a roughly apples-to-apples time series, we make the following two adjustments. First, from 2014 onwards, we use the 2013 ratio of U.S. to total market data revenue (72%) to get a U.S. market data revenue estimate. Second, from 2013 onwards, we subtract out 2012 tape revenue of $117M to get to U.S. market data revenue excluding tape plan revenue.[7] These two assumptions together imply that any revenue growth in Nasdaq's total market data category since 2014 is attributed 72% to Nasdaq's U.S. proprietary market data segment with the remaining 28% to international and index market data revenues. Our sense is that this convention is conservative since Nasdaq reports that both international market data revenue and index data revenue were relatively flat in the years leading up to 2013 (International: $83M in 2011 to $77M in 2013; Index: $24M in 2011 to $27M in 2013). We find that the Nasdaq reporting categories containing proprietary market data saw annual revenue growth of 13.7% for the period 2006-2012. We estimate that Nasdaq proprietary market data growth was 6.8% for the period 2012-2017, with the caveat that this 2012-2017 growth rate is based on some assumptions whereas the 2006-2012 growth rate is based off of numbers directly in Nasdaq filings.

We use four data sources for BATS co-location/connectivity revenue: BATS's 2012 S-1 statement (revenue from 2009-2011), BATS's 2016 S-1 statement (revenue from 2010-2015), the CBOE/BATS 2016 proxy statement (revenue for 9 months of 2016, which we annualize), and CBOE's 2017 annual report (which reports BATS's contribution to CBOE revenues for 10 months, which we annualize; CBOE's acquisition of BATS was finalized on Feb 28 2017). BATS states in its 2012 S-1 that it began charging for co-location/connectivity revenue, described as "port fees," in Q4 2009 (pg. 64) so we report numbers starting in 2010. Before 2012, the reporting segment was called "Other Revenues" in BATS's 2012 S-1; BATS describes the category by stating "Other revenues consist of port fees, which represent fees paid for connectivity to our markets, and, more recently, additional value-added products revenues." The reporting segment changed to "Port Fees and Other" in BATS's 2016 S-1; the revenue reported for 2011 in BATS's 2016 S-1 is within 1% of the 2011 revenue reported in BATS's 2012 S-1 so we conclude that the change was almost entirely a renaming of the reporting category rather than a substantive change. The reporting segment changed again to "Connectivity Fees and Other" for 2016 in the CBOE/BATS proxy statement, which was a change in name only (revenue reported from previous years are consistent with the 2016 S-1). In 2017, as a part of CBOE, BATS's revenue from co-location/connectivity is split across two CBOE segments, "Access fees" and "Exchange services and other fees." CBOE separately reports BATS's contributions to these categories and we report their sum, annualized to twelve months. For BATS co-location and connectivity, revenue more than quadrupled (growth of 64.0% per year) from 2010 through 2013, the last full year before BATS's acquisition of Direct Edge (which combined EDGX and EDGA under the same exchange company as BZX and BYX). Revenue then doubled from 2013 to 2015, but likely in large part due to the Direct Edge acquisition in 2014, and then grew 11.7% per year from 2015 to 2017.

BATS reports in its 2016 S-1 that its two original BATS exchanges, BZY and BYX, only began charging for proprietary market data in Q3 2014 (pg. 94). We thus use numbers starting in 2015 (this is also the first full year that revenue associated with Direct Edge is included in BATS's filings). BATS market data revenue for U.S. equities is included in the category "Market Data Fees" in BATS's 2016 S-1 (2015 revenue) and in the 2016 CBOE/BATS proxy statement (9 months of revenue from 2016, which we annualize). BATS's market data revenue for 2017 comes from CBOE's 2017 annual report, which provides BATS's contribution to the category

---

[7] We feel comfortable treating tape revenues as flat since 2012 since tape revenues are based on depth and market shares, and we show in Appendix C that market shares are roughly flat and show in Appendix F.2 that depth shares line up one-for-one with market shares. Nasdaq's market shares by share volume from 2011 to 2017 were: 29.1%, 29.5%, 28.9%, 31.3%, 28.0%, 26.0%, 28.0%. Moreover, in the last three years that Nasdaq did report tape revenues separately, they were essentially constant: $117M in 2010, $115M in 2011, and $117M in 2012.

"Market Data Fees" for 10 months in 2017 (which we annualize). We know that tape revenue is a significant fraction of market data revenue, and utilize percentages provided in BATS's filings to estimate tape revenue, which we can subtract from overall market data revenue as reported in BATS's filings. BATS reports in its 2016 S-1 that 84% of market data revenue in 2015 comes from tape revenue (pg. 21), and reports in the CBOE/BATS merger proxy that 79% comes from tape revenue in 2016 (pg. 56). Using these percentages, we estimate that 2015 tape revenue is \$110M and 2016 tape revenue is \$116M. We also assume that 2017 tape revenue is flat from 2016 levels.[8] If we subtract these tape revenue estimates from the overall market data revenue reported in BATS's filings, we get \$21.0M in 2015, \$30.8M in 2016 and \$38.3M in 2017 (growth of 35.3% per year). These numbers include data revenues related to BATS's European equities and U.S. options business, so they overstate the level but likely understate the growth rate of BATS's U.S. proprietary market data since 2015.[9]

Overall the data, while imperfect, are suggestive of exchanges "discovering a new pot of gold" in the Reg NMS era—that is, discovering that they could charge significant money for something they used to not charge for. To summarize, the overall annual growth rates are: 15.9% for Nasdaq co-location/connectivity (2006-2017), 10.5% for Nasdaq proprietary market data (2006-2017), and 40.4% for BATS co-location/connectivity (2010-2017). If we use 10% as a conservative overall growth rate for ESST revenue since 2015, and apply this growth rate to our estimates of revenue from Section 2.2, this implies that 2018 ESST revenues are between \$899M-\$1,053M.

---

[8] BATS's combined market share for its four exchanges by share volume declined slightly from 36.3% in 2016 to 34.5% in 2017, so if anything 2017 tape revenues would be slightly smaller than 2016.

[9] The CBOE/BATS proxy statement reports on pg. 311-312 that, of the growth in market data revenue of \$10.7M for the first 9 months of 2016 versus the same period in 2015, \$7.4M came from U.S. proprietary market data ("pricing changes in proprietary market data that were implemented in the third quarter of 2015 and the first quarter of 2016") versus \$0.7M from U.S. options (pg. 312) and \$0.7M from European equities (pg. 318). Unfortunately, we are not able to ascertain any specific information about growth in BATS's proprietary market data revenues from 2016 onwards; once BATS became incorporated into CBOE reporting became even less granular.
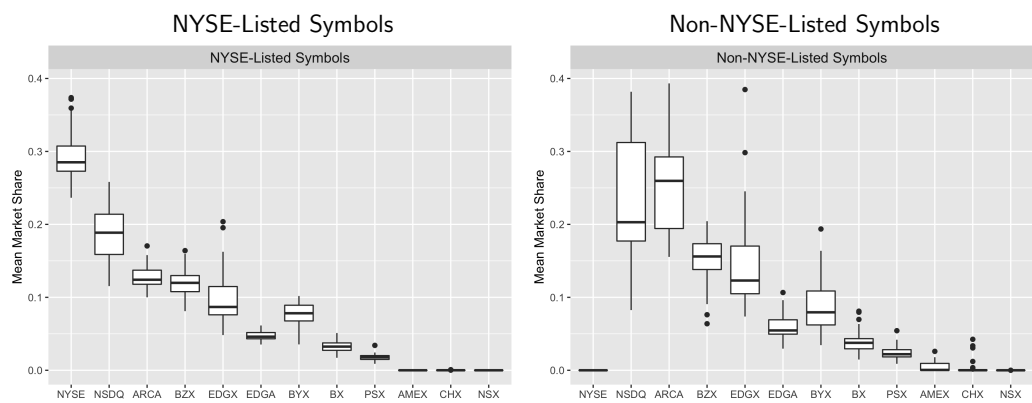
Figure C.1: Exchange Market Shares, 2011 - 2015
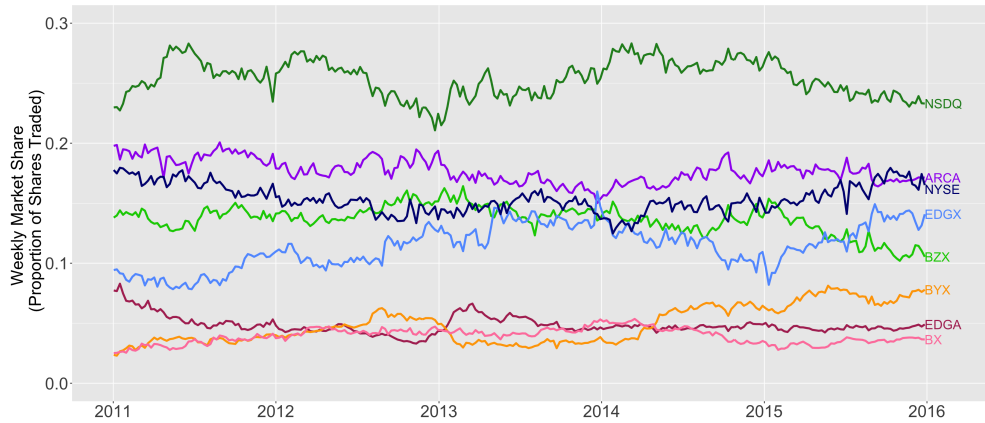
Notes: The data is from NYSE TAQ and covers January 2011 to Dec 2015 for the Top 8 exchanges. The market shares are based on all on-exchange trading volume in shares.

C   Supporting Details for Exchange Market Shares

Figure C.1 depicts exchange market shares from 2011 to 2015 for the Top 8 exchanges (which handle 98% of regular-hours trading volume). The figure shows that market shares are clearly interior and relatively stable over time. Of the top 8 exchanges with meaningful market share, the highest among them is roughly 25%.

Figure C.2 explores how market shares vary across stocks. For each stock in the top 100, we compute its average market share per exchange over all dates in 2015. We then present this data as a box plot. Each box represents the 25th-75th percentile range for symbol market shares on that exchange, with the solid horizontal line in the middle of the box representing the median. The lines above and below the box represent the full range, with dots for outliers. As can be seen, while there is of course variation across symbols, most of the variation in the data is driven by the exchange.

Figure C.2: 2015 Exchange Market Shares: Per Stock

Notes: The data is from NYSE TAQ. Observations are symbol-exchange averages of symbol-exchange-date market shares in 2015. In a given box, the middle line is the median and the edges of the box are the 25th and 75th percentiles. The lines on top of and below the box (whiskers) go out to the interquartile range multiplied by +/-1.5. The dots are symbol-exchange outliers that fall outside of that range.

Figure C.3: Exchange Market Shares: 2007 – 2015

Reg NMS Era Weekly Market Shares
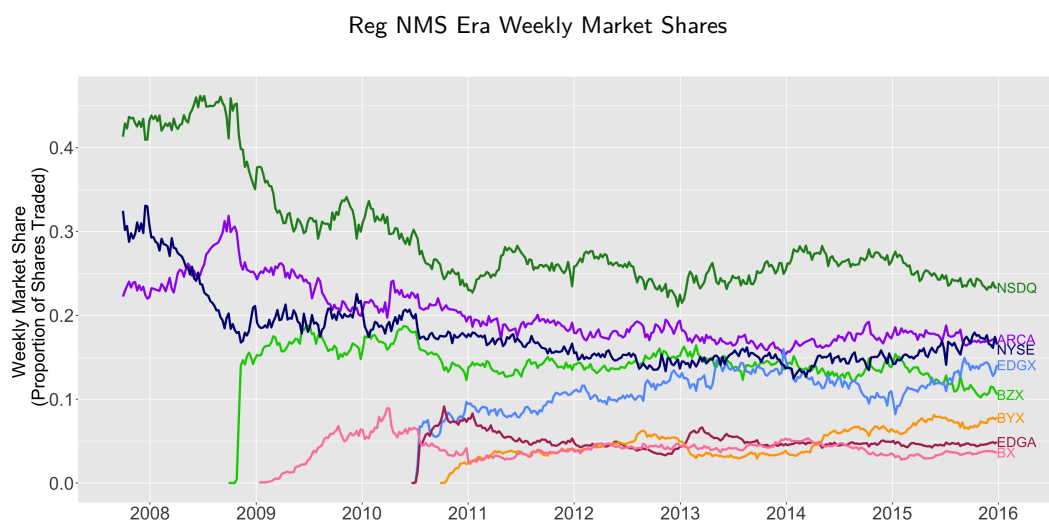
Notes: The data is from NYSE TAQ and covers October 2007 to December 2015 for the Top 8 exchanges. The market shares are based on all on-exchange trading volume in shares.

We also extend the timeline of market shares. Figure C.3 presents exchange market shares for the Top 8 exchanges from October 2007, the start of the Reg NMS era, through the end of 2015. There are several "jumps" in the data, in particular for BZX in late 2008 and EDGA and EDGX in 2010. Although these exchanges were officially approved at the time of the jump, they operated as off-exchange venues, or ATS's, and had significant market shares before they were officially approved. Thus, although the data show jumps in market share when these exchanges were approved, the market share change in going from an ATS to an exchange was likely much more smooth.

D   Institutional Background: UTP and Reg NMS

This appendix provides further details regarding Unlisted Trading Privileges (UTP) and Regulation National Market System (Reg NMS), the two key U.S. stock market regulations which guide our model and which were discussed more briefly in Section 2.4.

We note that while our discussion focuses on the United States, there are economically similar regulations for stock exchanges in Canada and somewhat similar regulations in Europe.[10] Regulations for futures exchanges, on the other hand, are quite different from those for stock exchanges, both in the U.S. and abroad. In particular, there is no analogue of UTP in futures markets because each contract is proprietary to a particular exchange. Similarly, there are differences between the regulation of stock exchanges and the regulation of financial exchanges for other financial instruments like government bonds, corporate bonds, foreign currency, etc.; in particular, the information dissemination provisions of Reg NMS are often economically different in these asset classes.

D.1   Unlisted Trading Privileges (UTP)

Section 12(f) of the 1934 Exchange Act, passed by Congress, directed the Securities and Exchange Commission to "make a study of trading in unlisted securities upon exchanges and to report the results of its study and its recommendations to Congress." Since that time, the right of one exchange to facilitate trading in securities that are listed on other exchanges has undergone several evolutions. In its current form, passed by Congress in the Unlisted Trading Privileges Act of 1994 (U.S. Congress, 1994) and clarified by the SEC in a Final Rule effective November 2000 (U.S. Securities and Exchange Commission, 2000), one exchange may extend unlisted trading privileges (UTP) to a security listed on another exchange immediately upon the security's initial public offering on the listing exchange, without any formal application or approval process through the SEC. Prior to 1994, exchanges had to formally apply to the SEC for the right to extend UTP to a particular security; such approval was "virtually automatic" following a delay of about 30-45 days (Hasbrouck, Sofianos and Sosebee, 1993). Between the passage of the UTP Act of 1994 and the Final Rule in 2000, extension of UTP was automatic but only after an initially two-day, and then one-day, delay period after the security first began trading on its listing exchange. For further historical discussion of UTP, please see the background section of the 2000 Final Rule document, and also Amihud and Mendelson (1996).

D.2   Regulation National Market System (Reg NMS)

Regulation National Market System ("Reg NMS"), was passed in June 2005 and implemented beginning in October 2007 (U.S. Securities and Exchange Commission, 2005). It is a long and complex piece of regulation, with roots tracing to the Securities Exchange Act Amendments of 1975 and the SEC's "Order Handling Rules"

---

[10]In both Canada and Europe, stocks are fungible across venues as in the US. In Canada's version of the Order Protection Rule (which goes by the same name), the key difference is that the rule applies to the full depth of the order book, not just the first level (Canadian Securities Administrators, 2009). In Europe, instead of the (prescriptive) Order Protection Rule there are (principles-based) best execution regulations (Petrella, 2010). Note however that principles-based best execution requirements leave some ambiguity with regard to whether market participants have to "pay attention" to quotes from small exchanges, which could affect innovation incentives; whereas under the Order Protection Rule there is no such ambiguity. This seems a good topic for future research.

promulgated in 1996.[11] For the purpose of the present paper, however, there are two core features to highlight.[12]

The first is the Order Protection Rule, or Rule 611. The Order Protection Rule prohibits an exchange from executing a trade at a price that is inferior to that of a "protected quote" on another exchange.[13] A quote on a particular exchange is "protected" if it is (i) at that exchange's current best bid or offer; and (ii) "immediately and automatically accessible" by other exchanges. Reg NMS does not provide a precise definition of "immediately and automatically accessible," but the phrase certainly included automated electronic continuous limit order book markets and certainly excluded the NYSE floor system with human brokers. A June 2016 rules clarification issued by the SEC, in conjunction with its approval of IEX's symmetric speed bump (see Appendix H for details), indicated that exchanges can use market designs that impose delays on the processing of orders and still qualify as "immediate and automatic" so long as (i) the delay is of a *de minimis* level of less than 1 millisecond, and (ii) the purpose of the delay is consistent with the efficiency and fairness goals of the 1934 Exchange Act (U.S. Securities and Exchange Commission, 2016*c*). This rules clarification suggests that quotes on an exchange that adopts frequent batch auctions would be protected under Rule 611, so long as the batch interval satisfies the de minimis delay standard, though to date the SEC has not further clarified its position. Also ambiguous is whether an asymmetric speed bump would be permitted under this rules clarification, though there, the SEC has explicitly signaled some hesitation, on the basis that the asymmetry is unfair; please see further discussion of the Chicago Stock Exchange and Cboe EDGA asymmetric speed bump proposals in Appendix H, and please see MacKenzie (2021) for a broader context on the debates about asymmetric speed bumps. Budish (2016*c*) encourages the SEC to proactively clarify what market designs are and are not allowed within the de minimis rule.

The second key provision to highlight is the Access Rule, or Rule 610. Intuitively, in order to comply with the Order Protection Rule, exchanges and market participants must be able to efficiently obtain the necessary information about quotes on other exchanges and efficiently trade against them. As the SEC writes (pg. 26), "...protecting the best displayed prices against trade-throughs would be futile if broker-dealers and trading centers were unable to access those prices fairly and efficiently."

The Access Rule has three sets of provisions that together are aimed at ensuring such efficient "search and access." First, Rule 610(c) limits the trading fee that any exchange can charge to 0.3 pennies, which, importantly, is less than the minimum tick size of 1 penny. This ensures that if one exchange has a strictly better displayed price than another exchange, the price is economically better after accounting for fees. Second, Rule 610(d) has provisions that together ensure that prices across markets do not become "locked" or "crossed" — specifically, each exchange is required to monitor data from all other exchanges and to ensure that it does not display a quote that creates a market that is locked (i.e., bid on one exchange equal to ask on another exchange) or crossed

---

[11] The goal of the National Market System is described by the SEC as follows: "The NMS is premised on promoting fair competition among individual markets, while at the same time assuring that all of these markets are linked together, through facilities and rules, in a unified system that promotes interaction among the orders of buyers and sellers in a particular NMS stock. The NMS thereby incorporates two distinct types of competition—competition among individual markets and competition among individual orders—that together contribute to efficient markets." See U.S. Securities and Exchange Commission (2005), pg 12.

[12] For an overview of Reg NMS, a good source is the introductory section of the SEC's final ruling itself (U.S. Securities and Exchange Commission, 2005). For an overview of the National Market System prior to Reg NMS, good sources are O'Hara and Macey (1997) and the SEC's "Market 2000" study (U.S. Securities and Exchange Commission, 1994).

[13] As noted in the main text, sophisticated market participants can take on responsibility for compliance with the Order Protection Rule themselves, absolving exchanges of the responsibility for checking quotes on other exchanges, by using an order type denoted intermarket sweep order (ISO). The relevant aspects of Reg NMS are Rule 600(b)(30) for the definition of ISOs, Rule 611(b)(5) for the exchange's exemption from ensuring compliance with the Order Protection Rule for ISOs, and Rule 611(c) for this compliance obligation instead residing in the sender of the ISO.

(i.e., bid on one exchange strictly greater than an ask on another exchange). Together, then, rules 610(c) and 610(d) ensure that there is a well-defined "national best bid and offer" (NBBO) across all exchanges (at least ignoring the complexities that arise due to latency, see Section 4 of Budish, 2016b). Third, Rule 610(a) prevents exchanges from charging discriminatory per-share trading fees based on whether the trader in question does or does not have a direct relationship with the exchange. In our model, the notion of a direct relationship with the exchange is captured by the decision of whether to buy exchange-specific speed technology, which represents exchange products like proprietary data feeds, co-location, and connectivity. What Rule 610(a) ensures is that market participants face the same trading fee schedule, whether or not they have such a direct relationship.

E Theory Appendix

E.1 Preliminaries

Trading Firms. In this appendix, we index all TFs by i or k and any references to a particular TF i or k without explicitly noting its speed will allow for that TF to be either fast or slow.

Investor Strategies. In the equilibria that we construct, investors employ *routing table strategies* in period 2 of each trading game to break ties when they are indifferent over transacting on different exchanges. Routing table strategies are defined to be a vector of fixed weights $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_M)$ such that, in the event that there are multiple exchanges contained in set $\mathcal{J}$ that offer depth at the same best price net of trading fees, an investor determines how much to purchase on each exchange using the following procedure: the investor "consumes" liquidity at rate $\gamma_j / \sum_{k \in \mathcal{J}} \gamma_k$ on each exchange $j \in \mathcal{J}$ until either (i) its demand is satisfied, or (ii) one or more exchanges in $\mathcal{J}$ no longer has any depth remaining at the best price, in which case the investor updates the set of exchanges that offer depth at the best price, updates its rate of consumption across exchanges, and continues consuming liquidity as before.[14]

Orders and Liquidity Provision. As defined in the main text, we denote by $\mu_{ij} \in \mathcal{S}$ the *message set* for TF i submitted to exchange j, where $\mathcal{S}$ is the set of all potential combinations of messages. TFs can send three types of messages to each exchange j: (i) standard limit orders, which take the form (q_i, p_i) and indicate that the TF is willing to buy (if $q_i > 0$) or sell (if $q_i < 0$) up to $|q_i|$ units at price p_i; (ii) cancellations of existing limit orders in $\boldsymbol{\omega}_j$, i.e., exchange j's order book; and (iii) immediate-or-cancel orders (IOCs), which are standard limit orders that, if not fully executed in a given period, have any portion that is remaining cancelled by the exchange at the end of the period. A message set submitted to a particular exchange may also contain no messages (i.e., $\mu_{ij} = \emptyset$); such a message set *maintains* the TF's existing limit orders on that exchange, if any exist.[15] A TF can *adjust* an existing limit order (e.g., change the price) by sending a set of messages cancelling the old limit order and providing a new one. Denote by $\boldsymbol{\mu}_i \equiv \{\mu_{ij}\}_{j \in \mathcal{M}}$ the message sets submitted by TF i to all exchanges, where $\mathcal{M}$ represents the set of all exchanges.

---

[14] For example, consider three exchanges across which $(0.25, 1, 1)$ units of the security are available at the same best price net trading fees. An investor with unit demand and routing table strategies $(1/3, 1/3, 1/3)$ initially consumes liquidity at an equal rate from all three exchanges until the first exchange no longer has any available liquidity at the best price; this occurs after the investor has purchased 0.75 units in total. For the last 0.25 units that the investor demands, the investor consumes equally from the remaining two exchanges. To operationalize this strategy, the investor submits a limit order to purchase $(0.25, 0.375, 0.375)$ units from the three exchanges. Note that routing table strategies allow investors to (essentially) employ lexicographic preferences over exchanges: e.g., if there are three exchanges with depth available at the best price net trading fees, allowing $\boldsymbol{\gamma} = (1 - \varepsilon - \varepsilon^2, \varepsilon, \varepsilon^2)$ for $\varepsilon > 0$ sufficiently small approximates an investor consuming depth from exchange 1 before moving to exchange 2, and then consuming depth from exchange 2 before moving to exchange 3. If $\gamma_j = 0$ for all $j \in \mathcal{J}$, we assume that an investor splits his demand uniformly among exchanges contained in $\mathcal{J}$. It is without loss to assume that routing table strategy weights sum to 1.

[15] For thinking practically about equilibria in our Stage 4 trading game, it is important to emphasize that our use of the language "submit messages" in this Appendix will include the possibility that a message set contains no messages to one or more exchanges (i.e., $\mu_{ij} = \emptyset$ for TF i and some exchanges j). As noted in the main text, trading games can be interpreted as lasting a sufficiently short amount of time (e.g., one millisecond or potentially even less) so that in most trading games, no exogenous Period-2 events occur. In the equilibria that we analyze, in this case that no exogenous Period-2 events occur, then in the next Period 1 all TFs simply maintain their existing limit orders in $\boldsymbol{\omega}$ by utilizing these "empty" message sets. It is in this sense that these equilibria will capture the idea that each exchange's limit order book settles into a rest point between Period-2 events, i.e., between arrivals of an investor, informed trader, or public information.

We introduce the following terms for liquidity provision in this Appendix. We say that a *limit order provides liquidity* if it offers to buy (or sell) some positive quantity at a price less than (or greater than) the current value of y and within the support of J. Denote by $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$ the set of TF i's liquidity-providing limit orders on exchange j, given the prior state of exchange j's order book $\boldsymbol{\omega}_j$ and the processing of any messages contained in μ_{ij}. We say that a *message set μ_{ij} provides liquidity* on exchange j if $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$ is non-empty; this can occur if either (i) μ_{ij} contains a limit order that provides liquidity; or (ii) TF i has outstanding liquidity-providing limit orders on exchange j and μ_{ij} does not cancel all of these limit orders. We say μ_{ij} *provides l units of liquidity at spread s* if $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$ contains a limit order to buy the security at $y - s/2$, and a limit order to sell at $y + s/2$ for some (bid-ask) spread $s \geq 0$.

There are also two sets of relationships between message sets that we refer to in this Appendix. We say that $\boldsymbol{\mu}'_i$ (weakly) *withdraws liquidity* relative to $\boldsymbol{\mu}_i$ if any limit order providing liquidity (at a given price and quantity on a particular exchange) contained in $\boldsymbol{\mu}'_i$ is also contained in $\boldsymbol{\mu}_i$, and any messages contained in $\boldsymbol{\mu}'_i$ but not in $\boldsymbol{\mu}_i$ are cancellations of existing limit orders. This implies that, for every exchange j, any limit order providing liquidity contained in $LO_{ij}(\mu'_{ij}; \boldsymbol{\omega}_j)$ is also contained in $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$. We say that $\boldsymbol{\mu}'_i$ is a (strict) *price improvement* over $\boldsymbol{\mu}_i$ if, for any $q \in (0, 1]$ and any exchange $j \in \mathcal{M}$, buying and selling q units on exchange j is weakly cheaper trading against limit orders in $LO_{ij}(\mu'_{ij}; \boldsymbol{\omega}_j)$ than against limit orders in $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$, and there exists some quantity $q \in (0, 1]$ and exchange j for which it is strictly cheaper to buy or sell q units trading against limit orders in $LO_{ij}(\mu'_{ij}; \boldsymbol{\omega}_j)$ than against limit orders in $LO_{ij}(\mu_{ij}; \boldsymbol{\omega}_j)$. (If there is no ability to buy (or sell) q units trading against limit orders in $LO_{ij}(\cdot)$ on exchange j, then the cost of buying (or selling) q units against limit orders in $LO_{ij}(\cdot)$ is considered infinite.) Note that if $\boldsymbol{\mu}_i$ does not provide liquidity on any exchange, then any $\boldsymbol{\mu}'_i$ providing liquidity at any (finite) price on any exchange represents a price improvement over $\boldsymbol{\mu}_i$.

E.2 Order Book Equilibrium (OBE)

Let $E\pi_i(\boldsymbol{\mu}_i, \boldsymbol{\mu}_{-i})$ represent TF i's expected profits from a trading game given the Period-1 message sets $\boldsymbol{\mu}_i$ that it submits and the Period-1 message sets submitted by all other trading firms, denoted $\boldsymbol{\mu}_{-i} \equiv \{\mu_{kj}\}_{k \neq i, j \in \mathcal{M}}$, taking as given the state $(y, \boldsymbol{\omega})$ at the beginning of Period 1 of the given trading game and anticipated Period 2 behavior described in the main text. Expectations are taken over the potentially random sequence in which the message sets $(\boldsymbol{\mu}_i, \boldsymbol{\mu}_{-i})$ are processed by exchanges in Period 1, the random action of nature in Period 2, and, in the event of a sniping race in Period 2 on one or more Continuous exchanges, the random sequence in which TFs' message sets are processed by exchanges.

We restate the OBE definition from the main text with additional formalism below:

Definition E.1. An *order book equilibrium* (abbreviated OBE) of our trading game is a set of message sets $\boldsymbol{\mu}^* \equiv \{\boldsymbol{\mu}^*_i\}$ submitted by all TFs in Period 1 given state $(y, \boldsymbol{\omega})$ that satisfies the following two conditions:

1. *No safe profitable price improvements.* No TF i has a strictly profitable price improvement that is *safe,* defined as remaining strictly profitable even if some other TF profitably withdraws liquidity in response to TF i's deviation.

   Formally, for any TF i, if $\boldsymbol{\mu}'_i$ is a price improvement over $\boldsymbol{\mu}^*_i$ and is strictly profitable meaning that $E\pi_i(\boldsymbol{\mu}'_i, \boldsymbol{\mu}^*_{-i}) > E\pi_i(\boldsymbol{\mu}^*_i, \boldsymbol{\mu}^*_{-i})$, then there is some other TF k and profitable reaction $\boldsymbol{\mu}'_k$ that withdraws liquidity relative to $\boldsymbol{\mu}^*_k$ and renders TF i's deviation no longer strictly profitable: i.e.,

$E\pi_i(\boldsymbol{\mu}'_i, (\boldsymbol{\mu}'_k, \boldsymbol{\mu}^*_{-ik})) \leq E\pi_i(\boldsymbol{\mu}^*_i, \boldsymbol{\mu}^*_{-i})$ and $E\pi_k(\boldsymbol{\mu}'_k, (\boldsymbol{\mu}'_i, \boldsymbol{\mu}^*_{-ik}))) > E\pi_k(\boldsymbol{\mu}^*_k, (\boldsymbol{\mu}'_i, \boldsymbol{\mu}^*_{-ik}))$, where $\boldsymbol{\mu}^*_{-ik}$ denotes message sets in $\boldsymbol{\mu}^*$ for all TFs other than TF i and TF k.

2. *No robust deviations.* No TF i has any other strictly profitable deviation (i.e., not a price improvement) that is *robust*, defined as remaining strictly profitable if, in response to TF i's deviation, some other TF engages in a profitable reaction that is either: (a) a withdrawal of liquidity; or (b) a safe profitable price improvement (as defined in 1.).

Formally, for any TF i, if $E\pi_i(\boldsymbol{\mu}'_i, \boldsymbol{\mu}^*_{-i}) > E\pi_i(\boldsymbol{\mu}^*_i, \boldsymbol{\mu}^*_{-i})$ for some deviation $\boldsymbol{\mu}'_i$ that is not a price improvement over $\boldsymbol{\mu}^*_i$, then there is some other TF k and profitable reaction $\boldsymbol{\mu}'_k$ that renders TF i's deviation no longer strictly profitable, and either: (a) $\boldsymbol{\mu}'_k$ withdraws liquidity relative to $\boldsymbol{\mu}^*_k$; or (b) $\boldsymbol{\mu}'_k$ is a safe profitable price improvement, and hence is a profitable price improvement that remains strictly profitable for TF k even if any other TF, including TF i, withdraws liquidity in response.

Discussion and Example. As discussed in the main text, OBE strictly weakens Markov perfect equilibrium (MPE) for our repeated trading game (and Nash equilibrium for a single play of our trading game) by relaxing the requirement that no strictly profitable unilateral deviations exist for any set of Period-1 orders. OBE instead only requires that there are no strictly profitable unilateral deviations *that remain strictly profitable* even if another TF were to profitably react in a particular way.

To provide intuition for how Conditions 1 and 2 help ensure both the existence of equilibria and the uniqueness of certain equilibrium features, consider a single play of the Stage 4 trading game with a single Continuous exchange that charges zero trading fees and all N fast TFs have purchased ESST from this exchange. In this setting, assume that the trading game begins with an empty limit order book where $y = 10$ and $s^*_{continuous} = 2$.

First, note that there does not exist an MPE or OBE of this trading game where only one unit of liquidity is provided at a spread strictly greater than $s^*_{continuous}$ in Period 1. To see why, assume that in Period 1, some TF i provides one unit of liquidity at spread $s = 4$ around y (i.e., bid 8, ask 12), and no other liquidity is provided by any other TF. This violates Condition 1, as there exists a profitable deviation that is a safe profitable price improvement for some fast TF $k \neq i$ to provide a unit of liquidity at spread $s' = 4 - 2\varepsilon$ around y (i.e., bid $8 + \varepsilon$, ask $12 - \varepsilon$) for any $\varepsilon \in (0, 1]$. This price improvement by TF k is also safe because even if TF i reacted to the deviation by withdrawing his liquidity, TF k strictly prefers providing liquidity at spread s' to sniping TF i's liquidity at spread $s > s^*_{continuous}$.

Next, note that there does not exist an MPE or OBE where only one unit of liquidity is provided at a spread strictly narrower than $s^*_{continuous}$ in Period 1. To see why, assume that in Period 1, some TF i provides one unit of liquidity at spread $s = 1$ around y (i.e., bid 9.5, ask 10.5), and no other liquidity is provided by any other TF. This violates Condition 2, as there exists a profitable deviation that is also robust for TF i to widen its spread to $s^*_{continuous}$ around y (i.e., bid 9, ask 11) on its one unit of liquidity. This deviation is robust as there is no other liquidity that could be profitably withdrawn, and there is no safe profitable price improvement for any other TF (as fast TFs earn more from sniping liquidity at $s^*_{continuous}$ than providing liquidity at a strictly narrower spread, and slow TFs earn strictly negative profits providing liquidity at a strictly narrower spread).

Thus, both MPE and OBE rule out equilibria in which one unit of liquidity is provided at any spread other than $s^*_{continuous}$: at strictly wider spreads, there are safe profitable price improvements to offer liquidity at more

competitive prices; at strictly narrower spreads, there are robust deviations to increase prices on the offered liquidity.

Now consider a candidate equilibrium where TF i provides one unit of liquidity at exactly $s^*_{continuous}$ around y (i.e., bid 9, ask 11), and no other liquidity is provided by any other TF. As we discuss below, there exist three types of strictly profitable unilateral deviations, implying that this candidate equilibrium is not an MPE. However, none of these deviations remain strictly profitable if other TFs are able to engage in the reactions discussed above, and hence these deviations do not prevent this candidate equilibrium from satisfying the conditions for OBE.

First, there are strictly profitable price improvements that involve some fast TF $k \neq i$ adding a unit of liquidity at bid $9 + \varepsilon$, ask $11 - \varepsilon$ for sufficiently small $\varepsilon > 0$: in doing so, TF k attempts to "have his cake and eat it too" by earning revenues from liquidity provision at a strictly narrower spread while also continuing to snipe TF i's orders that it just undercut. However, these deviations are not safe (Condition 1): TF i can profitably withdraw liquidity in response to TF k's price improvement and render the deviation unprofitable, as TF k would prefer to snipe i's liquidity at $s^*_{continuous}$ than provide liquidity at a narrower spread.

Second, consider the deviations by some fast TF $k \neq i$ to provide additional liquidity $l > 0$ at spread $s^*_{continuous}$. Such deviations are strictly profitable for TF k for $l \leq 1$ when TF i is slow, in a variation of the "have your cake and eat it too" deviation: when TF i is slow, TF k's liquidity is added to the order book prior to any of TF i's liquidity, and TF k can earn both revenues from liquidity provision and revenues from sniping TF i. However, these deviations, which are not price improvements, are not robust as there is a profitable withdrawal for TF i that renders them no longer strictly profitable (Condition 2a): following a withdrawal by TF i, TF k would earn the same amount from deviating as it would have from not deviating and sniping TF i's single unit of liquidity at $s^*_{continuous}$.

Last, there are strictly profitable deviations for TF i to widen its spread — for example, to say $s' = 4$ (i.e., bid 8, ask 12) — on its one unit of liquidity. However, these deviations are not robust as there is a safe profitable price improvement reaction by some fast TF $k \neq i$ that renders them unprofitable (Condition 2b): for example, if TF i deviates by widening its spread to $s' = 4$, then TF k can profitably react by providing a unit of liquidity at bid $8 + \varepsilon$, ask $12 - \varepsilon$ for sufficiently small $\varepsilon > 0$. This profitable price improvement remains strictly profitable for k even if TF i were to withdraw any liquidity, and thus is safe. (This is the same reasoning used above to show that a single unit of liquidity provided at any spread $s > s^*_{continuous}$ is not an equilibrium.)

Hence, the deviations that challenge the existence of an MPE for our repeated trading game (or a Nash equilibrium for a single play of our trading game), no longer do so for an OBE.

E.3   Proofs for Section 3.4 ("The Status Quo")Supporting Lemmas for Proposition 3.1

The proof of Proposition 3.1 relies on the following supporting lemmas.

Lemma E.1. *Consider the Stage 4 trading game with all Continuous exchanges, where: (i) all N fast TFs have purchased ESST from the same set of exchanges; and (ii) trading fees are zero for all exchanges contained in the non-empty set $\mathcal{J} \subseteq \mathcal{M}$ and strictly positive for all exchanges $m \notin \mathcal{J}$. Then:*

1. Existence: for any vector of market shares $\boldsymbol{\sigma^*} = (\sigma^*_1, \ldots, \sigma^*_M)$ such that $\sum_{j \in \mathcal{J}} \sigma^*_j = 1$ and $\sigma^*_m = 0$ if $m \notin \mathcal{J}$, there exists an OBE in which TFs in aggregate provide σ^*_j units of liquidity on each exchange j at spread $s^*_{continuous}$ in Period 1.

2. Uniqueness: any OBE has exactly one unit of liquidity provided in aggregate at spread $s^*_{continuous}$ in Period 1, where liquidity is provided across exchanges according to some vector of market shares $\boldsymbol{\sigma^*} = (\sigma^*_1, \ldots, \sigma^*_M)$ such that $\sum_{j \in \mathcal{J}} \sigma^*_j = 1$ and $\sigma^*_m = 0$ if $m \notin \mathcal{J}$.

(We do not require the uniqueness portion of Lemma E.1 for our main results, but state and prove it here for completeness.)

Proof. Condition on state (y, ω) at the beginning of this trading game.

Existence. Consider any vector of exchange market shares $\boldsymbol{\sigma^*} = (\sigma^*_1, \ldots, \sigma^*_M)$ such that $\sum_{j \in \mathcal{J}} \sigma^*_j = 1$ and $\sigma^*_m = 0$ if $m \notin \mathcal{J}$. Consider the following candidate equilibrium strategies. In Period 1, a single TF i (either fast or slow) submits messages to each exchange $j \in \mathcal{J}$ to provide exactly σ^*_j units of liquidity at spread $s^*_{continuous}$ around y (maintaining, adjusting or withdrawing any outstanding liquidity from the previous trading game as necessary). All other TFs do not provide any liquidity (which includes withdrawing any existing liquidity in ω, if present). In Period 2, an investor sends IOCs to all exchanges that, in aggregate, trade up to one unit in their desired direction, prioritizing their demand across exchanges based on the lowest value of $s_j/2 + f_j$ (where for each exchange j, s_j is the lowest spread at which liquidity is offered and f_j is the trading fee), and breaking ties according to routing table strategies given by $\boldsymbol{\gamma^*} = \boldsymbol{\sigma^*}$; additionally, if there are any remaining orders that are profitable to trade against based on the publicly observed state y, the investor trades against those as well. An informed trader sends IOCs to trade against any orders on any exchange that are profitable to trade against based on their privately observed y. If there is a publicly observable jump in y, TF i sends messages to cancel all liquidity providing orders; and all fast TFs not providing liquidity send IOCs to each exchange j to try to trade against (snipe) any orders that are profitable to trade against based on the new value of y.

We now show that the period-1 strategies comprise an OBE. First consider deviations by TF i. It is not profitable for TF i to adjust the quantity of liquidity that it provides: withdrawing any amount of liquidity offered at spread $s^*_{continuous}$ on any exchange is not strictly profitable; and offering additional liquidity beyond the initial one unit on any exchange is strictly unprofitable, as doing so only incurs additional adverse selection and sniping costs without any additional benefits. Reducing the spread on any amount of liquidity on any exchange is also strictly unprofitable. Last, although there is a strictly profitable deviation by TF i to increase its spread to $s' > s^*_{continuous}$ for $l \leq 1$ units of liquidity that it provides across some set of exchanges, such a deviation is not robust. To see why, consider as a reaction the profitable price improvement by some fast TF $k \neq i$ to provide l units at spread $s^*_{continuous}$ on the same set of exchanges, and an additional $1 - l$ units as a *stub quote* (i.e., liquidity provided at a spread outside the support of J). This reaction renders TF i's deviation unprofitable; furthermore, the reaction is safe since k would prefer to offer such liquidity even if TF i were to withdraw any of its liquidity: providing l units of liquidity at $s^*_{continuous}$ is strictly preferable to sniping the same amount of liquidity at $s' > s^*_{continuous}$, and the stub quote ensures that k prefers to engage in its reaction even if TF i were to withdraw any of its liquidity. Hence, there are no robust deviations (or safe profitable price improvements) for TF i.

Next, consider potential deviations for other TFs (who do not provide any liquidity given equilibrium strategies):

1. Consider the deviation by some TF $k \neq i$ to provide $l > 0$ additional units of liquidity at some spread $s' > s^*_{continuous}$ on any set of exchanges. This is strictly unprofitable for both slow and fast TFs, as the additional liquidity incurs only adverse selection and sniping costs without any benefits of being traded against by an investor.

2. Consider the deviation by some TF $k \neq i$ to provide $l > 0$ additional units of liquidity at spread $s' = s^*_{continuous} - \varepsilon$ for $\varepsilon > 0$ on any set of exchanges. If TF k is slow, this deviation is strictly unprofitable as slow TFs earn negative profits in expectation when offering any liquidity at a spread strictly less than $s^*_{continuous}$. If TF k is fast, this "undercutting" of TF i is strictly profitable for sufficiently small $\varepsilon > 0$ as TF k earns revenues from both liquidity provision (earning priority over i at a cost of just ε) and from sniping TF i's liquidity. But this deviation, which is a profitable price improvement for TF k, does not remain strictly profitable (and hence is not safe) if TF i profitably withdraws l units of its own liquidity offered at spread $s^*_{continuous}$ in response: by (3.1), liquidity provision and stale quote sniping are equally profitable at $s^*_{continuous}$ for a fast TF, implying that TF k would have preferred to snipe at $s^*_{continuous}$ than provide liquidity at a strictly narrower spread, $s' < s^*_{continuous}$.

3. Consider the deviation by some TF $k \neq i$ to provide $l > 0$ additional units of liquidity at $s^*_{continuous}$ on any set of exchanges. If TF k is slow, this deviation is not strictly profitable as slow TFs do not earn strictly positive expected profits when offering liquidity at spread $s^*_{continuous}$. If TF k is fast, this deviation is only strictly profitable for some $l > 0$ if (i) strictly less than one unit of liquidity is resting in the order book from the previous trading game (at spread $s^*_{continuous}$), and (ii) TF i is slow. To see why (i) is required for the deviation to be profitable, note that if one unit (or more) of liquidity is resting in the order book from the previous trading game, then any additional liquidity provided by TF k in Period 1 would have worse queue priority than this unit, and thus would not be filled by an investor upon arrival in Period 2. TF k's liquidity would then only bear sniping and adverse selection costs, and earn negative profits. To see why (ii) is required for the deviation to be profitable, note that if TF i is fast and less than one unit of liquidity is resting in the order book from the previous trading game, TF k's liquidity for sufficiently small $l > 0$ will be added to the order book in Period 1 before any of TF i's new liquidity with probability $1/2$ (due to the random sequence in which orders are processed by the exchange among TFs with the same speed technology) and thus filled by an investor upon arrival in Period 2. TF k's deviation would then earn in expectation:

$$l \times \left(\frac{1}{2} \cdot \lambda_{invest} \cdot \frac{s^*_{continuous}}{2} - (\frac{N-1}{N}\lambda_{public} + \lambda_{private}) \cdot L(s^*_{continuous}) \right),$$

where the term $\frac{1}{2} \cdot \lambda_{invest}$ reflects the probability that TF k's liquidity is added before TF i's new liquidity and is filled by an investor in Period 2. Substituting in $\lambda_{invest} \cdot \frac{s^*}{2} = (\lambda_{public} + \lambda_{private}) \cdot L(s^*_{continuous})$ from (3.1) into the previous expression and simplifying yields:

$$l \times (\frac{2-N}{2N}\lambda_{public} - \frac{1}{2}\lambda_{private}) \cdot L(s^*_{continuous}),$$

which is strictly negative since $N \geq 3$. In the case that (i) and (ii) are both satisfied, then the deviation for sufficiently small $l > 0$ is strictly profitable for the same reason that the deviation discussed above involving liquidity provision at a strictly narrower spread $s' = s^*_{continuous} - \varepsilon$ (for sufficiently small $\varepsilon > 0$) is profitable: TF k earns revenues from both liquidity provision and from sniping slow TF i's liquidity. However in this case, also as above, TF i has a profitable reaction to withdraw l units of its own liquidity, rendering the deviation by TF k not strictly profitable and hence not robust.

Hence, there are no safe profitable price improvements or robust deviations for any TF $k \neq i$. Thus, these message sets comprise an OBE for Period 1 given state (y, ω).

Note that in this equilibrium, in each trading game each fast TF earns (gross ESST fees) expected profits of $\sigma_j^* \times \frac{\Pi_{continuous}^*}{N}$ on exchange j from either liquidity provision or sniping activity; this implies that each fast TF earns in aggregate $\frac{\Pi_{continuous}^*}{N}$ per-trading game across all exchanges whether it provides liquidity or snipes stale quotes. Further, slow TFs earn zero in expectation, whether they provide liquidity or do nothing. Hence, repeated play of this OBE also comprises an equilibrium of the repeated Stage 4 trading game.

Uniqueness. Consider any equilibrium where $\boldsymbol{l} = (l_1^*, \ldots, l_M^*)$ units of liquidity are provided across exchanges at the end of Period 1. In any OBE, we now prove that exactly a single unit of liquidity is provided in aggregate among all exchanges with zero trading fees (i.e., $\sum_{j \in \mathcal{J}} l_j^* = 1$ and $l_m^* = 0$ if $m \notin \mathcal{J}$) at spread $s_{continuous}^*$ around y following Period 1. This follows from establishing the following three results.

First, in Period 1, exactly one unit of liquidity must be provided in aggregate across all exchanges. Assume by contradiction that an OBE exists with $l > 1$ units of liquidity offered at the end of Period 1. Focus on liquidity offered at the worst price. If such liquidity would never be filled by an investor in Period 2—which can occur if there is at least one unit of liquidity offered at a strictly better price—then any TF offering such liquidity would have a robust deviation to withdraw this liquidity, as such liquidity only bears adverse selection and sniping costs without liquidity provision benefits; thus, this cannot be an OBE. Hence, if there are $l > 1$ units of liquidity offered, all liquidity offered at the worst price must be in expectation filled by an investor in Period 2 with some probability that is strictly positive, but less than 1 (since $l > 1$). However, in this case, any TF offering liquidity at the worst price has a profitable price improvement to reduce the spread on its liquidity by some small amount $\varepsilon > 0$, thereby ensuring that its liquidity would be filled by an investor with certainty in Period 2; furthermore, this deviation remains profitable even if other TFs withdrew liquidity, and hence is safe. Contradiction. Assume next by contradiction that an OBE exists with $l < 1$ units of liquidity offered at the end of Period 1. Consider the strictly profitable unilateral deviation by any fast TF to offer $1 - l$ additional units of liquidity at spread $s_{continuous}^*$ on any exchange in $\mathcal{J}$. This is a safe profitable price improvement, as reactions that withdraw offered liquidity do not render this deviation weakly unprofitable. This cannot be an OBE; contradiction. Hence, exactly a single unit of liquidity must be offered at the end of Period 1 in any OBE.

Second, all liquidity that is provided on any exchange $j \in \mathcal{J}$ must be provided at spread $s_{continuous}^*$. Assume by contradiction that there exists an OBE where exactly one unit of liquidity in aggregate is offered at the end of Period 1, but there exists some amount of liquidity on exchange $j \in \mathcal{J}$ that is not offered at spread $s_{continuous}^*$. Assume first that in such an equilibrium, $l \le 1$ units are offered at a spread $s < s_{continuous}^*$ by some TF i on exchange j. Consider the strictly profitable unilateral deviation by TF i to increase its spread to $s_{continuous}^*$ on its offered liquidity on exchange j. This deviation is robust: there is no withdrawal or safe profitable price improvement that renders the deviation weakly unprofitable, as any fast TF considering a price improvement that undercuts TF i would instead prefer to snipe TF i's liquidity at $s_{continuous}^*$ as opposed to providing liquidity at a narrower spread, and any slow TF cannot profitably provide liquidity at any spread $s' \le s_{continuous}^*$. This cannot be an OBE; contradiction. Assume next that in such an equilibrium, $l \le 1$ units are offered at a spread $s > s_{continuous}^*$ by some TF i on exchange j. There is a safe profitable price improvement by some fast TF $k \ne i$ to undercut and provide l units at spread $s_{continuous}^*$, as there are no withdrawals of liquidity that render the deviation weakly unprofitable (since TF k prefers to provide liquidity at $s_{continuous}^*$ to sniping liquidity provided at $s > s_{continuous}^*$). This cannot be an OBE; contradiction.

Third, any positive quantity of liquidity cannot be provided on any exchange m where $f_m > 0$. Assume not, and there exists an equilibrium in which some quantity of liquidity $l > 0$ is provided on exchange m at some spread s' by some (fast or slow) TF i. Consider first the case where $\frac{s'}{2} + f_m > \frac{s_{continuous}^*}{2}$. In this case, there

is a safe profitable price improvement for some fast TF $k \neq i$: TF k can profitably provide the same amount of liquidity l on any exchange $j \in \mathcal{J}$ at spread $s^*_{continuous}$ (since an investor would strictly prefer to transact on exchange j at spread $s^*_{continuous}$ than on exchange m at s'), and this would remain strictly profitable for TF k even if TF i were to withdraw its liquidity in response (as TF k would strictly prefer to provide liquidity on exchange j at $s^*_{continuous}$ than snipe liquidity on exchange m at s'). Contradiction. Consider next the case where $\frac{s'}{2} + f_m \leq \frac{s^*_{continuous}}{2}$. In this case there is a strictly profitable deviation for TF i, that is also robust, to withdraw all liquidity on m and offer the same amount of liquidity on any exchange $j \in \mathcal{J}$ at spread $s^*_{continuous}$ (and avoid trading fees). The reason this deviation is robust is that there are no withdrawals (since only i provides liquidity), and there are no safe profitable price improvements by other TFs that render the deviation not strictly profitable: any fast TF $k \neq i$ prefers sniping liquidity on j at $s^*_{continuous}$ than offering it on any exchange at a lower spread, and any slow TF cannot profitably provide liquidity at a lower spread. Contradiction.

Thus, any OBE involves exactly a single unit of liquidity provided in aggregate at spread $s^*_{continuous}$ around y following Period 1, and liquidity is only provided on exchanges with zero trading fees. Given Period-2 strategies, it then follows that transaction volume upon the arrival of an investor in Period 2 coincides with liquidity provision for all exchanges (i.e., $\sigma^*_j = l^*_j$ for all $j \in \mathcal{M}$).

Lemma E.2. *("Lone-Wolf Lemma") Consider the Stage 4 trading game in any subgame with only Continuous exchanges where: (i) trading fees on all exchanges are zero; (ii) fast TF i, referred to as the "lone-wolf," has purchased exchange-specific speed technology (ESST) only on exchanges contained in the set $\mathcal{J} \subset \mathcal{M}$; and (iii) all other fast TFs have purchased ESST on the same set of exchanges $\mathcal{J}'$, where $\mathcal{J} \subset \mathcal{J}' \subseteq \mathcal{M}$. There exists an OBE for Period 1 of this trading game where exactly one unit of liquidity is provided only on exchanges contained in $\mathcal{J}$ by TF i at spread $\tilde{s}_N$ in Period 1, where $\tilde{s}_N$ solves:*

$$\lambda_{invest} \frac{\tilde{s}_N}{2} - \left(\frac{N-2}{N-1}\lambda_{public} + \lambda_{private}\right)L(\tilde{s}_N) = \frac{\lambda_{public}L(\tilde{s}_N)}{N} , \qquad (E.1)$$

and TF i earns in expectation at least $\pi_N^{lone-wolf} = \frac{N-1}{N^2}\lambda_{public}L(\tilde{s}_N)$ per-trading game gross of ESST fees, where $\pi_N^{lone-wolf} \in \left(\frac{N-2}{N-1} \times \frac{\Pi^*_{continuous}}{N}, \frac{\Pi^*_{continuous}}{N}\right)$. Furthermore, among OBE in which TF i is the sole liquidity provider in Period 1 of each trading game (i.e., all other TFs play strategies in which they provide no liquidity), it is unique that TF i provides exactly one unit of liquidity at spread $\tilde{s}_N$ solely on exchanges contained in $\mathcal{J}$.

Proof. In this proof, all references to TF profits are in expectation for each trading game, gross ESST fees.

Preliminaries. Define the spread $\bar{s}_N$ to be the minimum spread the lone-wolf TF i must charge on exchange j for one unit of liquidity so that i breaks even in expectation when the $N-1$ other fast TFs have also purchased ESST from j and no liquidity is provided on any other exchange; i.e., $\bar{s}_N$ is the solution to:

$$\lambda_{invest} \frac{\bar{s}_N}{2} - \left(\frac{N-1}{N}\lambda_{public} + \lambda_{private}\right)L(\bar{s}_N) = 0 . \qquad (E.2)$$

We refer to $\bar{s}_N$ as the zero-variable profit spread. The difference between the definition of $\bar{s}_N$ and the definition of $s^*_{continuous}$ in (3.1) is that $\bar{s}_N$ does not incorporate the opportunity cost of sniping, worth $\frac{1}{N}\lambda_{public}L(\cdot)$.

Next, we provide intuition for the spread $\tilde{s}_N$ defined in (E.1). Assume that conditions (i)-(iii) in the statement of the Lemma hold, and the lone-wolf TF i provides one unit of liquidity at spread $\tilde{s}_N$ on some exchange $j \in \mathcal{J}$. Then any other fast TF $k \neq i$ would be indifferent between (i) sniping TF i on exchange j (earning the right-hand side of (E.1)), and (ii) TF i not providing any liquidity, and TF k instead providing

one unit of liquidity at $\tilde{s}_N$ on some exchange $j' \notin \mathcal{J}$ (earning the left-hand side of (E.1), where TF k only risks being sniped by $N-2$ other TFs who have ESST on exchange j').

We now prove that $\bar{s}_N < \tilde{s}_N < s^*_{continuous}$. The first inequality, $\bar{s}_N < \tilde{s}_N$, follows from comparing (E.2) to (E.1), which can be re-written as $\lambda_{invest}\frac{\tilde{s}_N}{2} - ((\frac{1}{N} + \frac{N-2}{N-1})\lambda_{public} + \lambda_{private})L(\tilde{s}_N) = 0$. It is straightforward to show that the coefficient on λ_{public} is greater in (E.1) than in (E.2): $\frac{1}{N} + \frac{N-2}{N-1} = 1 - \frac{1}{N(N-1)} > 1 - \frac{1}{N} = \frac{N-1}{N}$. Hence, it follows that $\tilde{s}_N > \bar{s}_N$. The second inequality, $\tilde{s}_N < s^*_{continuous}$, follows using similar logic: in (3.1), which defines $s^*_{continuous}$, λ_{public} enters the equation with a coefficient of 1; however, in (E.1), which defines $\tilde{s}_N$, λ_{public} enters with a coefficient strictly less than 1.

The rest of the proof proceeds in three parts. First, we establish that an OBE with the properties outlined in the statement of the Lemma exists (Existence). Second, we establish that any OBE in which TF i is the sole liquidity provider in Period 1 of each trading game must have these properties (Uniqueness). Last, we prove that $\pi_N^{lone-wolf} \in (\frac{(N-2)}{(N-1)\times N} \times \Pi^*_{continuous}, \frac{\Pi^*_{continuous}}{N})$ (Profit Bound).

Existence. We now prove that there is an OBE for Period 1 of each trading game in which the lone-wolf TF i provides one unit of liquidity at spread $\tilde{s}_N$ across exchanges according to any arbitrary vector of shares $\boldsymbol{\sigma^*} = (\sigma^*_1, \ldots, \sigma^*_M)$ s.t. $\sum_{j \in \mathcal{J}} \sigma^*_j = 1$ and $\sigma^*_j = 0$ if $j \notin \mathcal{J}$, and no additional liquidity is provided by any other TF. Consider equilibrium strategies where in Period 1, TF i submits messages to provide one unit of liquidity at spread $\tilde{s}_N$ across exchanges in $\mathcal{J}$ according to $\boldsymbol{\sigma^*}$ (maintaining, adjusting or withdrawing any outstanding liquidity from the previous trading game as necessary), and other TFs do not provide any liquidity (which includes withdrawing any existing liquidity in ω, if present); and in Period 2, strategies follow those described in the proof of Lemma E.1 (where investors break ties across exchanges using routing table strategies $\boldsymbol{\gamma^*} = \boldsymbol{\sigma^*}$). The right-hand-side of (E.1) represents the gross expected payoffs that any fast TF $k \neq i$ expects to obtain by sniping TF i across all exchanges; the left-hand-side represents the gross expected payoffs that any fast TF k would anticipate if TF k were instead the sole liquidity provider on some other exchange $m \notin \mathcal{J}$ at spread $\tilde{s}_N$. Hence, no fast TF $k \neq i$ has a strictly profitable deviation—for example, by undercutting i or providing additional liquidity at a spread weakly smaller than $\tilde{s}_N$ on any exchange—that remains profitable if TF i reacts by withdrawing any liquidity that is no longer profitable to offer. Since $\tilde{s}_N < s^*_{continuous}$, any slow TF also has no robust deviations or safe profitable price improvements. Last, by similar arguments used in the proof of Proposition 3.1, there are no robust deviations for TF i: if TF i widened its spread on any amount of liquidity, the deviation would be rendered unprofitable by another fast TF k's safe profitable price improvement to provide that amount of liquidity on some exchange $m \notin \mathcal{J}$ at spread $\tilde{s}_N$ (which, by (E.1), is more profitable for TF k than sniping TF i at any spread strictly greater than $\tilde{s}_N$); and TF i reducing its spread or adjusting the amount of liquidity that it provides would strictly reduce profits. Thus, these strategies comprise an OBE.

Uniqueness. We now prove that among OBE in which TF i is the sole liquidity provider in Period 1 of each trading game, it is unique that TF i provides exactly one unit of liquidity at spread $\tilde{s}_N$ solely on exchanges contained in $\mathcal{J}$.

First, note that TF i must offer exactly one unit of liquidity in aggregate: otherwise, TF i would find it profitable to withdraw liquidity (if it offered strictly greater than one unit of liquidity) or have a safe profitable price improvement to add liquidity at spread $\tilde{s}_N$ on some exchange in $\mathcal{J}$ (if it offered strictly less than one unit of liquidity).

Second, note that such liquidity must be offered only on exchanges in $\mathcal{J}$. Assume not, and some positive amount of liquidity $l_m > 0$ is offered by TF i on some exchange $m \notin \mathcal{J}$. Consider first the case where such liquidity is offered by TF i at spread $s' > s^*_{continuous}$. In this case, there would be a safe profitable price

improvement by some other fast TF $k \neq i$ to undercut TF i and offer this amount of liquidity on exchange m at spread $s^*_{continuous}$; contradiction. Consider next the case where such liquidity is offered by TF i at spread $s' \leq s^*_{continuous}$. In this case, TF i would then have a robust deviation to withdraw that liquidity from m and offer instead the same amount of liquidity on some exchange in $\mathcal{J}$ at spread $\tilde{s}_N$ (as discussed above when establishing Existence, no fast TF $k \neq i$ would find offering liquidity at any spread less than $\tilde{s}_N$ on any exchange strictly preferable to sniping TF i's liquidity on an exchange in $\mathcal{J}$ at spread $\tilde{s}_N$); contradiction.

Last, TF i must offer this single unit of liquidity at spread $\tilde{s}_N$. If any amount of liquidity were offered at a lower spread, TF i would have a robust deviation to increase its spread to $\tilde{s}_N$; and if any amount of liquidity were offered at a strictly greater spread, there would be a safe profitable price improvement by some fast TF $k \neq i$ to provide the same amount of liquidity on some exchange $m \notin \mathcal{J}$ at spread $\tilde{s}_N$.

Profit Bound. Define

$$\pi^{lone-wolf}_N \equiv \lambda_{invest} \frac{\tilde{s}_N}{2} - (\frac{N-1}{N}\lambda_{public} + \lambda_{private})L(\tilde{s}_N) \tag{E.3}$$

to be the expected profits per trading game (gross ESST fees) that the lone-wolf TF i makes providing a single unit of liquidity at spread $\tilde{s}_N$ across exchanges contained in $\mathcal{J}$ when there are N total fast TFs (including him) that also have purchased ESST on exchanges contained in $\mathcal{J}$ and no other TF provides liquidity.

We now prove the stated bounds on $\pi^{lone-wolf}_N$. First, the upper bound,

$$\pi^{lone-wolf}_N < \frac{\Pi^*_{continuous}}{N} \equiv \lambda_{invest} \frac{s^*_{continuous}}{2} - (\frac{N-1}{N}\lambda_{public} + \lambda_{private})L(s^*_{continuous}) \,,$$

follows since, comparing the right-hand side of (E.3) to the right-hand side of the above expression, $\tilde{s}_N < s^*_{continuous}$ and $L(\tilde{s}_N) > L(s^*_{continuous})$. To obtain the lower bound, first solve for $\lambda_{invest} \frac{\tilde{s}_N}{2}$ in (E.1) and substitute this expression into the right-hand side of (E.3) to obtain:

$$\begin{aligned}
\pi^{lone-wolf}_N &= (\frac{1}{N} + \frac{N-2}{N-1} - \frac{N-1}{N})\lambda_{public}L(\tilde{s}_N) \\
&= \frac{N-2}{(N-1)N}\lambda_{public}L(\tilde{s}_N) \\
&> \frac{N-2}{(N-1)N}\lambda_{public}L(s^*_{continuous}) = \frac{N-2}{(N-1)N}\Pi^*_{continuous} \,,
\end{aligned}$$

where the inequality on the last line follows from $L(\tilde{s}_N) > L(s^*_{continuous})$.

E.3.1   Restriction on Negative Trading Fees

Before proceeding with our results for the full exchange competition game, we make the following remark regarding negative trading fees.

Remark E.1. Suppose some exchange $m \in \mathcal{M}$ sets a trading fee $f_m < 0$. For any arbitrarily large dollar amount $P > 0$, there exist TF strategies in Period 1 of any trading game in Stage 3 where TFs earn profits weakly greater than P, at the expense of exchange m, without engaging in self-dealing.

To see this, consider the following Period 1 strategies for any two distinct TFs i and k: TF i submits messages to exchange m to buy $Q \equiv P/(2 \times |f_m|)$ units at price y, and TF k submits messages to exchange m to sell Q units at price y. There are two cases to examine. First, if there is no other liquidity provided on exchange

m buying (or selling) at prices weakly greater than (or less than) y in Period 1, then TF i and k transact Q units with one another in Period 1, and TF i and k together earn $2 \times (Q \times |f_m|) = P$ in profits at the expense of exchange m. Second, if (for some reason) there is other liquidity provided on exchange m buying (or selling) at prices weakly greater than (or less than) y in Period 1, then it still must be the case that either TF i or TF k transacts at least Q units at a price no worse than y in that Period, implying that TFs trading these Q units, in aggregate, earn P in profits at the expense of exchange m.

Hence, if an exchange charges negative trading fees, it enables TFs, even without self-dealing, to take advantage and expose the exchange to arbitrarily large losses. For this reason, we restrict exchanges to charge non-negative trading fees.

E.3.2   Proof of Proposition 3.1 (Equilibrium of the Status Quo)

For any vector of ESST fees $\boldsymbol{F}^*$ that satisfies (3.2) and market shares $\boldsymbol{\sigma}^*$ such that $\sum_{j \in \mathcal{M}} \sigma_j^* = 1$, consider the following candidate equilibrium strategies following Stage 1 with only Continuous exchanges:

- In Stage 2, each exchange j charges F_j^* for ESST and sets trading fees $f_j = 0$;

- In Stage 3, all N fast TFs buy ESST from exchange j only if (i) its ESST fee $F_j \leq F_j^*$, (ii) $f_j = \min_{k \in M} f_k$, and (iii) $\sigma_j^* > 0$;

- In Stage 4, in Period 1 of each trading game:

  1. On the candidate equilibrium path: If all fast TFs purchase ESST from the same set of exchanges $\mathcal{J} \subseteq \mathcal{M}$ where $f_j = 0$ for all $j \in \mathcal{J}$, then in Period 1 of each trading game, some (fast or slow) TF i submits messages to provide $\sigma_j^*/(\sum_{k \in \mathcal{J}} \sigma_k^*)$ amount of liquidity on each exchange $j \in \mathcal{J}$ at spread $s_{continuous}^*$ around y (maintaining, adjusting or withdrawing any outstanding liquidity from the previous trading game as necessary), all other TFs submit messages such that they provide no liquidity on any exchange, and no liquidity is provided elsewhere.

  2. If one fast TF i purchases ESST from a non-empty strict subset of exchanges $\mathcal{J}' \subset \mathcal{M}$, and all other fast TFs $k \neq i$ purchase ESST from a strictly greater set of exchanges $\mathcal{J}$ (so that $\mathcal{J}' \subset \mathcal{J} \subseteq \mathcal{M}$) where $f_j = 0$ for all $j \in \mathcal{J}$, then in Period 1 of each trading game, TF i is the "lone-wolf" liquidity provider and submits messages to provide one unit of liquidity on some exchange $j \in \mathcal{J}'$ at spread $\tilde{s}_N$ (defined in (E.1)) around y (maintaining, adjusting or withdrawing any outstanding liquidity from the previous trading game as necessary), all other TFs submit messages such that they provide no liquidity on any exchange, and no liquidity is provided elsewhere.

  3. If one fast TF i purchases ESST from a set of exchanges $\mathcal{K} \subseteq \mathcal{M}$ and all other fast TFs $k \neq i$ purchase ESST from a strict subset of exchanges $\mathcal{J} \subset \mathcal{M}$, where $f_j = 0$ for all $j \in \mathcal{J}$ and $\mathcal{K} \nsubseteq \mathcal{J}$, then in Period 1 of each trading game:

     (a) If $\mathcal{J} \subset \mathcal{K}$ (so that TF i purchases from a strictly greater set of exchanges than all other TFs), strategies are as in Case 1 above and liquidity is provided only on exchanges in $\mathcal{J}$;

     (b) If $\mathcal{J} \cap \mathcal{K} = \emptyset$ (so that TF i purchases from no exchanges contained in $\mathcal{J}$), strategies are analogous to Case 1 above: some fast TF $k \neq i$ which has purchased ESST on exchanges in $\mathcal{J}$ submits messages to provide $\sigma_j^*/(\sum_{k \in \mathcal{J}} \sigma_k^*)$ amount of liquidity on each exchange $j \in \mathcal{J}$ at spread $s_{continuous}^*$ around y (maintaining, adjusting or withdrawing any outstanding liquidity

from the previous trading game as necessary) and all other TFs submit messages such that they provide no liquidity on any exchange;

(c) Otherwise (which occurs if $\mathcal{K}$ contains a non-empty strict subset of exchanges in $\mathcal{J}$ and at least one exchange outside of $\mathcal{J}$), strategies are as in Case 2 above where TF i is the lone-wolf liquidity provider, and provides one unit of liquidity at spread $\tilde{s}_N$ on some exchange contained in $\mathcal{J}' = \mathcal{J} \cap \mathcal{K}$ (maintaining, adjusting or withdrawing any outstanding liquidity from the previous trading game as necessary), all other TFs submit messages such that they provide no liquidity on any exchange, and no liquidity is provided elsewhere.

- In Stage 4, in Period 2 of each trading game, an investor sends IOCs to all exchanges that, in aggregate, trade up to one unit in their desired direction, prioritizing their demand across exchanges based on the lowest value of $s_j/2 + f_j$ (where for each exchange j, s_j is the lowest spread at which liquidity is offered and f_j is the trading fee), and breaking ties according to routing table strategies given by $\boldsymbol{\gamma}^* = \boldsymbol{\sigma}^*$; additionally, if there are any remaining orders that are profitable to trade against based on the publicly observed state y, the investor trades against those as well. An informed trader sends IOCs to trade against any orders on any exchange that are profitable to trade against based on their privately observed y. If there is a publicly observable jump in y, TF i sends messages to cancel all liquidity providing orders; and all fast TFs (except the liquidity provider TF i if i is fast) send IOCs to each exchange j to try to trade against (snipe) any orders that are profitable to trade against based on the new value of y.

Note that these candidate equilibrium strategies dictate play in all subgames that are reachable via any sequence of unilateral deviations in Stages 2 and 3. Specifically, Stage 3 strategies prescribes play for all fast TFs given any choice of ESST and trading fees chosen by exchanges in Stage 2. Stage-4-Period-1 strategies prescribe play in any subgame reachable following Stage 3 if no more than one fast TF engages in a deviation. For these Stage-4-Period-1 strategies, Case 1 prescribes play on the equilibrium path when no fast TF deviates, and all fast TFs purchase from the same set of exchanges. Cases 2 and 3(a)-(c) prescribe play when some fast TF i purchases from a different set of exchanges than all other fast TFs, and comprehensively covers play depending on whether TF i purchases ESST from a: (Case 2) strict subset, (Case 3(a)) strict superset, (Case 3(b)) non-overlapping set, or (Case 3(c)) partially-overlapping set of exchanges.

We now show that these strategies comprise an equilibrium.

First, consider Stage 2 deviations for exchanges regarding their choice of ESST fees and trading fees. If any exchange lowers its ESST fee, it strictly reduces its profits as it earns less from the sale of ESST but outcomes would otherwise remain the same. If any exchange increased its ESST fee, the exchange would earn zero profits as no fast TF would purchase ESST from it, and liquidity would only be provided on other exchanges in the subsequent OBE outcome of the Stage 4 trading game (see Lemma E.1). If any exchange increased its trading fee, it also would earn zero profits for the same reasons. Hence, exchanges have no strictly profitable unilateral deviations.

Next, we turn to Stage 3 strategies for fast TFs. By following candidate strategies in Stage 3 given exchanges did not deviate in Stage 2, all fast TFs earn $\frac{1}{N}\Pi^*_{continuous} - \sum_j F_j^*$ which, by condition (3.2), is positive. Potentially profitable unilateral deviations for any fast TF involve the purchase of ESST from a strict subset of exchanges (as purchasing ESST from no exchanges yields no profits, and being the only fast TF to purchase ESST from an exchange yields no benefit due to our fair-access assumption). In subgames following such deviations, prescribed strategies comprise the unique OBE for Period 1 of the subsequent "lone-wolf" Stage 4 trading game given no liquidity is provided by other TFs (Lemma E.2), and the deviating fast TF earns in

expectation $\pi_N^{lone-wolf}$ per trading game (gross trading fees). Condition (3.2) ensures that this deviation is not profitable for any fast TF. Similar arguments establish that there are no strictly profitable deviations for fast TFs in Stage 3 given at most one exchange engaged in any deviation in Stage 2.

Finally, given equilibrium play in Stages 2 and 3, Lemma E.1 establishes that Stage-4-Period-1 strategies comprise an OBE.

E.3.3   Proof of Proposition 3.2 (Bound on ESST Fees)

Consider any vector of ESST fees $\boldsymbol{F}' = (F'_1,...,F'_M)$ that maximizes $\sum_{j \in \mathcal{M}} F_j$ among all vectors of ESST fees that satisfy condition (3.2). This condition, satisfied by any equilibria described by Proposition 3.1, can be rewritten as:

$$\sum_{j:\sigma_j^*>0} F_j^* \le \frac{\Pi^*_{continuous}}{N} - \max(0, \pi_N^{lone\text{-}wolf} - \min_j F_j^*) \ .$$

Since the upper bound on the total sum of ESST fees across exchanges (the left-hand side) is increasing in the minimum ESST fee (on the right-hand side), such a vector $\boldsymbol{F}'$ must have the same ESST fees for all exchanges: i.e., there must be a constant $\tilde{F}$ such that $F'_j = \tilde{F}$ for all $j \in \mathcal{M}$. Hence, the vector of ESST fees that maximizes the sum over all ESST fees and satisfies condition (3.2) is unique and involves each exchange charging the same amount $\tilde{F}$. This implies that, in any equilibrium described by Proposition (3.1), each fast TF pays at most

$$M \times \tilde{F} \le \frac{1}{N}\Pi^*_{continuous} - (\pi_N^{lone\text{-}wolf} - \tilde{F})$$

in ESST fees across all exchanges. Substituting in the lower bound $\pi_N^{lone-wolf} > \frac{N-2}{(N-1)N}\Pi^*_{continuous}$ from Lemma E.2 into the above equation and re-arranging terms yields:

$$\tilde{F} < \frac{1}{(M-1)(N-1)N}\Pi^*_{continuous}$$
$$(\Leftrightarrow) \quad M \times N \times \tilde{F} < \frac{M}{(M-1)(N-1)}\Pi^*_{continuous} \ ,$$

where $M \times N \times \tilde{F}$ is the upper bound on the total amount of ESST fees earned by all exchanges.

E.4   Proofs For Sections 3.6-3.8

Preliminaries: Equilibrium Spreads on a Discrete Exchange.   Denote by $\bar{s}_{discrete}(f)$ the zero-variable profit spread for a liquidity provider on a Discrete exchange (henceforth, "Discrete") given Discrete charges a trading fee $f \ge 0$; such a spread solves:

$$\lambda_{invest}(\frac{\bar{s}_{discrete}(f)}{2} - f) - \lambda_{private}L(\bar{s}_{discrete}(f), f) = 0 \ , \tag{E.4}$$

where $L(s, f) \equiv E(J - \frac{s}{2} + f | J > \frac{s}{2} + f)Pr(J > \frac{s}{2} + f)$ represents the expected loss to a liquidity provider providing liquidity at spread s on an exchange with trading fee f in the event of being adversely traded against. The first term on the left-hand-side of (E.4) represents the revenues a liquidity provider earns when an investor arrives (i.e., half the spread less the trading fee), and the second term is the expected loss from informed trading. A unique solution $\bar{s}_{discrete}(f)$ exists for any $f \ge 0$ (and is strictly positive) by the same arguments used to establish the existence and uniqueness of $s^*_{continuous}$ in the main text (see the discussion following

equation (3.1)). Note that when the trading fee on Discrete is zero ($f = 0$), $\bar{s}(f) = s^*_{discrete}$, where $s^*_{discrete}$ is defined in (3.3) in the main text.

Define $f^*_{discrete}$ to be the trading fee so that an investor is indifferent between trading on Discrete at spread $\bar{s}_{discrete}(f^*_{discrete})$ with trading fee $f^*_{discrete}$, and trading on a Continuous exchange with no trading fee at the zero-variable profit spread $\bar{s}_{continuous} \equiv \bar{s}_N$, defined in equation (E.2). As the following lemma establishes, under a technical assumption, $f^*_{discrete}$ exists and is unique.

Lemma E.3. *Assume that the jump size distribution is continuously differentiable. Then there exists a unique solution $f^*_{discrete}$ to:*

$$\frac{\bar{s}_{discrete}(f^*_{discrete})}{2} + f^*_{discrete} = \frac{\bar{s}_{continuous}}{2} \ . \tag{E.5}$$

Furthermore, if $f < (>)f^*_{discrete}$, then $\frac{\bar{s}_{discrete}(f)}{2} + f < (>)\frac{\bar{s}_{continuous}}{2}$.

Proof. Let $H(s,f) = \lambda_{invest}(\frac{s}{2} - f) - \lambda_{private}L(s,f)$. Define $s(f)$ to be the solution to $H(s(f),f) = 0$ (hence, $\bar{s}_{discrete}(f) = s(f)$). Since the jump size distribution is continuously differentiable, $H(\cdot)$ is as well, and by the implicit function theorem (given $\frac{\partial H}{\partial s} \neq 0$, which is satisfied), the function $s(f)$ exists and is continuously differentiable with

$$s'(f) = -\frac{\frac{\partial H}{\partial f}}{\frac{\partial H}{\partial s}} = \frac{\lambda_{invest} + \lambda_{private}L_f(s,f)}{\lambda_{invest}/2 - \lambda_{private}L_s(s,f)} \ ,$$

where $L_f(\cdot)$ and $L_s(\cdot)$ represent partial derivatives of $L(\cdot)$. We next establish that $\frac{\bar{s}_{discrete}(f)}{2} + f$ is strictly increasing in f: differentiating this expression with respect to f implies that a sufficient condition for it to be strictly increasing in f is $s'(f) > -2$. Substituting in for $s'(f)$ and re-arranging terms yields:

$$\frac{\lambda_{invest}}{\lambda_{private}} > (L_s(s,f) - L_f(s,f)/2) \ .$$

This inequality always holds since the left-hand-side is strictly positive, and the right-hand-side is weakly negative.[16] Since $\frac{\bar{s}_{discrete}(f)}{2} + f$ is thus strictly increasing and continuous in f, and since it is less than $\frac{\bar{s}_{continuous}}{2}$ for $f = 0$ but greater than $\frac{\bar{s}_{continuous}}{2}$ when $f = \frac{\bar{s}_{continuous}}{2}$, there exists a unique solution to (E.5). The rest of the statement directly follows. $\qquad\square$

We maintain the assumption that the jump size distribution is continuously differentiable for the rest of this section.

---

[16] Let G_{jump} denote the jump size distribution and g_{jump} its associated density. To establish that $(L_s(s,f) - L_f(s,f)/2) \leq 0$, note that

$$L(s,f) = E(J - \frac{s}{2} + f|J > \frac{s}{2} + f)Pr(J > \frac{s}{2} + f) = \int_{\frac{s}{2}+f}^{\infty} [t - \frac{s}{2} + f]g_{jump}(t)dt \ .$$

Hence,

$$L_s(s,f) \ = \ -\int_{\frac{s}{2}+f}^{\infty} \frac{g_{jump}(t)}{2}dt - [(\frac{s}{2} + f) - \frac{s}{2} + f] \times \frac{g_{jump}(s/2 + f)}{2} = -\frac{(1 - G_{jump}(s/2+f))}{2} - f \times g_{jump}(\frac{s}{2} + f),$$

$$L_f(s,f) \ = \ \int_{\frac{s}{2}+f}^{\infty} g_{jump}(t)dt - [(\frac{s}{2} + f) - \frac{s}{2} + f] \times g_{jump}(\frac{s}{2} + f) = (1 - G_{jump}(\frac{s}{2} + f)) - 2f \times g_{jump}(\frac{s}{2} + f),$$

and $(L_s(s,f) - L_f(s,f)/2) = -(1 - G_{jump}(s/2 + f))$, which is weakly negative since $G_{jump}(x) \leq 1$ for all x.

E.4.1 Proof of Proposition 3.3 (Equilibrium of Stage 4 with a Single Discrete Exchange)

The statement of Proposition 3.3 assumes that all exchanges charge zero trading fees. Here, we prove a more general version of Proposition 3.3, and allow the single Discrete exchange to charge a weakly positive trading fee.

Consider any Stage 4 subgame with potentially multiple Continuous exchanges and single Discrete exchange (henceforth, "Discrete"), where all fast TFs have purchased ESST from the same set of Continuous exchanges, and trading fees are zero on all Continuous exchanges and equal to $\tilde{f} \in [0, f^*_{discrete})$ on Discrete (where $f^*_{discrete}$ is defined in (E.5)). We will prove that in any equilibrium of this Stage 4 subgame: exactly one unit of liquidity is provided on Discrete at bid-ask spread $\bar{s}_{discrete}(\tilde{f})$ around the current value of y following Period 1, and no liquidity is provided elsewhere; Period 2 behavior is as described in the statement of the Proposition; and that such an equilibrium exists.[17]

Existence. Consider the following Stage 4 strategies given state $(y, \boldsymbol{\omega})$. In Period 1, a single TF i submits messages to Discrete to provide one unit of liquidity at spread $\bar{s}_{discrete}(\tilde{f})$ around y; if he has liquidity outstanding from the previous trading game, he maintains, adjusts or withdraws it as necessary so that he provides exactly one unit at spread $\bar{s}_{discrete}(\tilde{f})$ around y on Discrete, and no liquidity on any Continuous exchange. All other TFs do not provide any liquidity on any exchange. In Period 2, an investor sends IOCs to all exchanges that, in aggregate, trade up to one unit in their desired direction, prioritizing their demand across exchanges indexed by j based on the lowest value of $s_j/2 + f_j$ (where for each exchange j, s_j is the lowest spread at which liquidity is offered), and breaking ties in an arbitrary fashion; additionally, if there are any remaining orders that are profitable to trade against based on the publicly observed state y, the investor trades against those as well. An informed trader sends IOCs to trade against any orders on any exchange that are profitable to trade against based on their privately observed y. If there is a publicly observable jump in y, TF i sends messages to cancel all liquidity providing orders; and all fast TFs not providing liquidity send IOCs to each exchange j to try to trade against (snipe) any orders that are profitable to trade against based on the new value of y.[18]

Using similar arguments used in the proof of Lemma E.1, it is straightforward to show that there are no safe profitable price improvements or robust deviations in Period 1, or profitable unilateral deviations in Period 2, and hence these strategies comprise an equilibrium for the Stage 4 trading game. For example, in Period 1, any increase by TF i in its spread on Discrete to $\bar{s}_{discrete}(\tilde{f}) + \varepsilon$ for any $\varepsilon > 0$ and any amount of liquidity is not a robust deviation, as it is rendered unprofitable by a safe profitable price improvement from another TF to provide the same amount of liquidity at spread $\bar{s}_{discrete}(\tilde{f}) + \varepsilon/2$ on Discrete.

Uniqueness. We now prove that in any Stage 4 equilibrium, exactly one unit of liquidity is provided on Discrete following Period 1 at spread $\bar{s}_{discrete}(\tilde{f})$, and no liquidity is provided on any Continuous exchange.

First, using similar arguments used in the proof of Lemma E.1, it is straightforward to establish that exactly one unit of liquidity in aggregate must be provided in any Stage 4 equilibrium.

Next, we show that some positive amount of liquidity cannot be provided on any Continuous exchange. Assume by contradiction that there exists an equilibrium in which some positive amount of liquidity is provided on some Continuous exchange at spread $\tilde{s}$. If $\tilde{s} < \bar{s}_{continuous}$ (the zero-variable profit spread on a Continuous exchange), then the liquidity provider would have a robust deviation to withdraw it. If $\tilde{s} > \bar{s}_{continuous}$, then

---

[17]It is straightforward to use the same arguments in this proof to establish that this statement holds even if there exist other Discrete exchanges with trading fees that are strictly greater than $\tilde{f}$.

[18]When $\tilde{f} = 0$, there also exist equilibria in which, following a publicly-observed jump in y that exceeds $s^*_{discrete}/2$, TF i does not withdraw all of its offered liquidity on Discrete, and multiple TFs submit IOCs to Discrete to purchase one unit of the security at price y. In such equilibria, all trades in the auction occur at the new value of y.

there is a safe profitable price improvement by any slow TF to provide the same amount of liquidity on Discrete at some spread $s' \in (\bar{s}_{discrete}(\tilde{f}), \bar{s}_{continuous} - 2\tilde{f})$. Contradiction.[19]

Last, we show that the one unit of liquidity on Discrete must be provided at spread $\bar{s}_{discrete}(\tilde{f})$. Assume by contradiction that there exists an equilibrium where some positive amount of liquidity on Discrete is provided at spread $\tilde{s} \neq \bar{s}_{discrete}(\tilde{f})$. If $\tilde{s} < \bar{s}_{discrete}(\tilde{f})$ (the zero-variable profit spread on Discrete given informed trading), then the liquidity provider would have a robust deviation to withdraw it. If $\tilde{s} > \bar{s}_{discrete}(\tilde{f})$, then there is a safe profitable price improvement by any slow TF to provide the same amount of liquidity at any spread $s' \in (\bar{s}_{discrete}(\tilde{f}), \tilde{s})$. Contradiction.

E.4.2 Proof of Proposition 3.4 (Equilibrium with a Single Discrete Exchange)

Existence. First, we establish that if a single Discrete exchange ("Discrete") charged any trading fee $f' > f^*_{discrete}$ (where $f^*_{discrete}$ is the solution to equation (E.5)) and some Continuous exchange m had zero trading fees, then in any Stage 4 equilibrium, no liquidity can be provided on Discrete. To see why, assume by contradiction that there exists an equilibrium in which Discrete charges a trading fee $f' > f^*_{discrete}$, and there is positive liquidity provided on Discrete in Stage 3. The lowest spread at which liquidity could be profitably offered on Discrete is the zero-variable profit spread $\bar{s}_{discrete}(f')$. At this spread, the total price considered by an investor contemplating trading on Discrete is $\bar{s}_{discrete}(f')/2 + f' > \bar{s}_{continuous}/2$ (by Lemma E.3 and the definition of $f^*_{discrete}$). This implies that there exists a safe profitable price improvement for some fast TF on Continuous exchange m to provide liquidity on exchange m at spread $s' \in (\bar{s}_{continuous}, \bar{s}_{discrete}(f') + 2f')$, as such liquidity at spread s' on exchange m would be preferred by investors to the liquidity on Discrete and earns strictly positive profits for the deviating TF. Contradiction.

Next, consider the following candidate equilibrium strategies. In Stage 2, Discrete charges positive trading fees $f^*_{discrete}$; all Continuous exchanges charge zero trading fees and zero ESST fees. In Stage 3, fast TFs do not purchase ESST from Continuous exchanges unless ESST fees are strictly negative. In Stage 4, market participants use strategies described in the Proof of Proposition 3.3, with the modification that investors break ties in favor of Discrete.[20]

We now show that these strategies comprise an equilibrium. In Stage 2, Continuous exchanges have no strictly profitable deviations: charging strictly positive ESST or trading fees do not affect profits; charging strictly negative ESST fees earns strictly negative profits. Discrete also has no strictly profitable deviations: by Proposition 3.3, reducing trading fees yields lower profits for Discrete as it does not affect Stage 4 trading game behavior; and any higher trading fees results in all trading activity in Stage 4 occurring on Continuous exchanges and zero profits (as established above). In Stage 3, there are no strictly profitable deviations by any fast TF (as purchasing ESST does not affect profits). Last, in Stage 4, similar arguments used in the Existence portion of the proof for Proposition 3.3 establish that these strategies comprise an equilibrium of the Stage 4 trading game.

Uniqueness. We now prove that in any equilibrium, (i) Discrete charges trading fees equal to $f^*_{discrete}$; (ii) in every iteration of the trading game, exactly one unit of liquidity is offered on Discrete at spread $\bar{s}_{discrete}(f^*_{discrete})$ and no liquidity is provided on any Continuous exchange; (iii) all Continuous exchanges earns zero profits; and

---

[19]As long as the spread on Discrete $s' < \bar{s}_{continuous} - 2\tilde{f}$, an investor would prefer to transact on Discrete (paying $\frac{s'}{2} + \tilde{f}$) than transact on Continuous (paying $\frac{\bar{s}_{continuous}}{2}$).

[20]If trading fees are restricted to be in discrete units (e.g., in units of $0.0001), then there also exist equilibria in which investors always break ties in favor of Continuous: in such equilibria, Discrete charges the greatest trading fee f such that $\frac{\bar{s}_{discrete}(\tilde{f})}{2} + f < \frac{\bar{s}_{continuous}}{2}$, and liquidity is only offered on Discrete in each trading game.

(iv) Discrete earns expected per-trading-game profits that exceed $\frac{N-1}{N}\Pi^*_{continuous}$.

For claim (i), first consider a candidate equilibrium where Discrete charges trading fee $f < f^*_{discrete}$. Since $\frac{\bar{s}_{discrete}(f)}{2} + f < \frac{\bar{s}_{continuous}}{2}$, then by continuity of $\bar{s}_{discrete}(\cdot)$, there exists trading fee $f' = f + \varepsilon$ for sufficiently small $\varepsilon > 0$ that Discrete could charge such that $\frac{\bar{s}_{discrete}(f')}{2} + f' < \frac{\bar{s}_{continuous}}{2}$ and would yield Discrete strictly higher profits as it would still capture all trading volume but obtain higher trading revenues; contradiction. Next, consider a candidate equilibrium where Discrete charges trading fee $f > f^*_{discrete}$. In this candidate equilibrium, either Discrete has zero trading volume in Stage 4, or positive trading volume. In the case that Discrete has zero trading volume, since there exists some strictly positive trading fee $f' < f^*_{discrete}$ that Discrete could charge such that $\frac{\bar{s}_{discrete}(f')}{2} + f' < \frac{\bar{s}_{continuous}}{2}$ and yields positive trading volume on Discrete in any Stage 4 equilibrium (by Proposition 3.3), there is a profitable deviation for Discrete to charge f' instead; contradiction. In the case that Discrete has positive trading volume, then using similar arguments as in Lemma E.3 and above, since there exists some strictly positive trading fee $f' > 0$ that a Continuous exchange m could charge such that $\frac{\bar{s}_{continuous}(f')}{2} + f' < \frac{\bar{s}_{discrete}(f)}{2} + f$ (where $\bar{s}_{continuous}(f')$ is the analogous zero-variable profit spread on a Continuous exchange given it charges trading fees f') and moves all trading volume to exchange m in any Stage 4 equilibrium, there is a profitable deviation for exchange m to charge f' instead; contradiction. Thus, Discrete cannot charge any trading fee $f > f^*_{discrete}$. Hence, Discrete must charge trading fees equal to $f^*_{discrete}$.

For claim (ii), any equilibria in which strictly less than or strictly greater than one unit of liquidity is provided in aggregate across all exchanges can be ruled out using similar arguments as in Lemma E.1. Now consider a candidate equilibrium in which exactly one unit of liquidity is offered in aggregate, but strictly positive liquidity is offered on some Continuous exchange. Then Discrete has a profitable deviation in Stage 2 by instead charging a trading fee $f' = f^*_{discrete} - \varepsilon$ for sufficiently small $\varepsilon > 0$, guaranteeing that all volume transacts on Discrete in Stage 4 and increasing profits for Discrete; contradiction.

Claim (iii) directly follows from (i) and (ii).

We have now proved that in any equilibrium, Discrete charges trading fees equal to $f^*_{discrete}$ and one unit of liquidity is provided in each trading game on Discrete at spread $\bar{s}_{discrete}(f^*_{discrete})$ and no liquidity is provided on any Continuous exchange. We now establish claim (iv).

First, substituting in the expression for $\frac{\bar{s}_{continuous}}{2} = \frac{\bar{s}_{discrete}(f^*_{discrete})}{2} + f^*_{discrete}$ given by (E.5) into (E.2) yields:

$$\lambda_{invest}\left(\frac{\bar{s}_{discrete}(f^*_{discrete})}{2} + f^*_{discrete}\right) - \left(\frac{N-1}{N}\lambda_{public} + \lambda_{private}\right)L(\bar{s}_{continuous}) = 0 . \tag{E.6}$$

Equation (E.4) provides the expression for the zero-variable profit spread on Discrete at $f^*_{discrete}$:

$$\lambda_{invest}\left(\frac{\bar{s}_{discrete}(f^*_{discrete})}{2} - f^*_{discrete}\right) - \lambda_{private}L(\bar{s}_{discrete}(f^*_{discrete}), f^*_{discrete}) = 0 .$$

Subtracting this expression from (E.6) and re-arranging yields:

$$2f^*_{discrete} \times \left(\lambda_{invest} + \lambda_{private}Pr(J > \frac{s^*_{discrete}(f^*_{discrete})}{2} + f^*_{discrete})\right) = \frac{N-1}{N}\lambda_{public}L(\bar{s}_{continuous}) , \tag{E.7}$$

(since $L(\bar{s}_{discrete}(f^*_{discrete}), f^*_{discrete}) - L(\bar{s}_{continuous}) = 2f^*_{discrete} \times Pr(J > \frac{s^*_{discrete}(f^*_{discrete})}{2} + f^*_{discrete})$). The left-hand side of (E.7) represents Discrete's total expected revenues per-trading game from trading fees $f^*_{discrete}$: Discrete earns $2f^*_{discrete}$ in trading fees each time an investor arrives and trades (with probability λ_{invest}), or

an informed trader arrives and trades (with probability $\lambda_{private} \times Pr(J > \frac{s^*_{discrete}(f^*_{discrete})}{2} + f^*_{discrete}))$. The right-hand side of (E.7) represents $\frac{(N-1)}{N}$ share of the total "sniping prize" at a spread of $\bar{s}_{continuous}$; since $\bar{s}_{continuous} < s^*_{continuous}$ and $L(\cdot)$ is decreasing in the spread, the right-hand side is strictly greater than $\frac{(N-1)}{N}$ share of $\Pi^*_{continuous}$, and the result follows.

E.4.3 Proof of Proposition 3.5 (Equilibrium with Multiple Discrete Exchanges)

Existence. Consider the following candidate equilibrium strategies. In Stage 2, all Discrete exchanges charge zero trading fees; any Continuous exchange charges zero trading fees and zero ESST fees. In Stage 3, fast TFs do not purchase ESST from any Continuous exchange with weakly positive ESST fees, and purchase ESST from any Continuous exchange with strictly negative ESST fees. In Stage 4 Period 1, a single TF i submits messages to any Discrete exchange with the minimum trading fee (denoted j) to provide one unit of liquidity at spread $\bar{s}_{discrete}(\tilde{f})$ around y; if he has liquidity outstanding from the previous trading game, he maintains, adjusts or withdraws it as necessary so that he provides exactly one unit at spread $\bar{s}_{discrete}(\tilde{f})$ around y on Discrete, and no liquidity on Continuous. All other TFs do not provide any liquidity on any exchange (which includes withdrawing any existing liquidity in ω, if present). In Period 2, investors, informed traders, and TFs use strategies described in the Proof of Proposition 3.3, with the modification that investors break ties in favor of Discrete exchange j.[21]

We now check that these strategies comprise an equilibrium. In Stage 2, Continuous exchanges have no profitable deviations: charging strictly positive ESST or trading fees do not affect profits; charging strictly negative ESST fees earns strictly negative profits. Any Discrete exchange also has no strictly profitable deviations: increasing trading fees results in no trading volume and revenues given equilibrium strategies. In Stage 3, there are no strictly profitable deviations by any fast TF (as purchasing ESST does not affect profits). Last, similar arguments used in the Existence portion of the proof for Proposition 3.3 establish that these strategies comprise an equilibrium of the Stage 4 subgame.

Uniqueness. We now prove that in any equilibrium, (i) at least one Discrete exchange charges zero trading fees; (ii) in every iteration of the trading game, exactly one unit of liquidity is offered only on Discrete exchanges with zero trading fees at spread $s^*_{discrete}$ (equivalent to $\bar{s}_{discrete}(0)$) around the current value of y following Period 1; (iii) no liquidity is provided on Discrete exchanges with positive trading fees or on Continuous exchanges; and (iv) all exchanges and trading firms earn zero profits. For claim (i), consider a candidate equilibrium where all Discrete exchanges charge strictly positive trading fees, and the minimum trading fee is $f > 0$. The same logic underlying why undifferentiated Bertrand competition results in marginal cost pricing implies that this cannot be an equilibrium: for some Discrete exchange, there exists a profitable Stage 1 deviation to charge a slightly lower trading fee $f' = f - \varepsilon$ for some $\varepsilon > 0$ as this would guarantee that all subsequent trading volume would occur on that exchange (the same arguments used in Proposition 3.3 establish that all Stage 4 equilibria involve all trading volume occurring on the Discrete exchange with the lowest trading fee). Claims (ii) and (iii) follow directly from the arguments used in Proposition 3.3. Claim (iv) directly follows from claims (i) and (ii).

---

[21]Denote by $\mathcal{J}$ the set of all Discrete exchanges with zero trading fees and let $\boldsymbol{\sigma}^*$ be any arbitrary vector of market shares. As in Proposition 3.1, there also exist equilibria with multiple Discrete exchanges in which TF i provides $\sigma^*_j / (\sum_{k \in \mathcal{J}} \sigma^*_k)$ amount of liquidity on each exchange $j \in \mathcal{J}$ (thereby providing one unit of liquidity in aggregate), and investors break ties when indifferent according to routing table strategies given by $\boldsymbol{\gamma}^* = \boldsymbol{\sigma}^*$.

E.4.4 Proof of Proposition 3.6 (Prisoner's Dilemma Payoffs)

The result follows directly from $\Pi^*_{discrete} > 0$ and the following Lemma.

Lemma E.4. *Assume that ESST fees $\boldsymbol{F}^*$ satisfy condition (3.2). Then $NF^*_j < \Pi^*_{discrete}$ for any exchange j.*

Proof. Let $\bar{F}$ be the most that any exchange j can charge for ESST fees given (3.2). This maximum is realized for exchange j when all other exchanges charge 0; condition (3.2) then becomes

$$\bar{F} \leq \frac{1}{N}\Pi^*_{continuous} - \pi^{lone-wolf}_N < \frac{1}{(N-1)N}\Pi^*_{continuous} \, ,$$

where the last inequality follows from the lower bound on $\pi^{lone-wolf}_N$ given by Lemma E.2. Hence, $N\bar{F} < \frac{1}{N-1}\Pi^*_{continuous}$. Since $\Pi^*_{discrete} > \frac{N-1}{N}\Pi^*_{continuous}$ by Proposition 3.4 (and since $N \geq 3$), the result follows. $\quad\square$

F   Additional Empirical Evidence on the Virtual Single Platform Theory

In this appendix, we report evidence supporting the Virtual Single Platform Theory presented in Section 3.5.1. We show that all active exchanges have the same equilibrium bid and offer, i.e., quoted prices are identical across exchanges, and each exchange's share of market depth (i.e., its share of liquidity) at this common best bid and offer equals its share of market volume.

Data.   We use the Daily NYSE Trade and Quote ("TAQ") dataset accessed via Wharton Research Data Services. The data contain every trade and every top-of-book quote update for every exchange, for all U.S. listed stocks and exchange-traded funds (ETFs), timestamped to the millisecond. The key advantage of this data, for our purposes, is that it is comprehensive across exchanges and labels every trade and quote update by exchange.

For the results presented in this section, we make three types of sample restrictions. First, we use data from all trading days in 2015.[22] Second, we limit our sample to the top 8 exchanges by market share, which constitute 98% of trading volume. We separately report results for the top 5 exchanges (which constitute 83% of trading volume) and the top 8. The top 5 exchanges all utilize what is commonly referred to as the "maker-taker" pricing model in which the taker of liquidity is charged a fee and the provider of liquidity is paid a rebate. The next 3 exchanges all use the "inverted" (or "taker-maker") pricing model in which the taker is paid a rebate and the maker pays a fee, and this difference in fee structure relative to the larger exchanges raises some subtleties for the analysis (which we discuss below). The remaining 4 exchanges active during 2015, sometimes called the "regional" exchanges, together had about 2% market share. Anecdotally, industry participants regard them as vestiges of an earlier era of stock exchange competition. Third, we restrict attention to the 100 most heavily traded stocks and ETFs. In 2015, there were 9,175 symbols that traded at least once; however, most trade relatively infrequently. We also require that the symbols in our sample satisfy a set of data-cleaning filters: trading continuously throughout the year under the same ticker, having a share price of at least $1, not having a listing change, and having at least $10 million in average daily volume. These 100 symbols constitute about one-third of daily volume.[23]

Taker-Maker.   The difference between taker-maker and maker-taker fee structures means that liquidity at the same displayed quoted price is economically more attractive for the taker of liquidity on the taker-maker exchange than on the maker-taker exchange. Put differently, if a liquidity provider offers liquidity at the same quoted price on both types of exchanges, it is in effect offering a better price on the taker-maker exchange. For example, suppose there is displayed depth on both a taker-maker exchange and a maker-taker exchange at the national best offer price, say $10.00. Suppose that on both exchanges, the rebate is $0.0029, the fee is $0.0030, and hence the net fee collected by the exchange (i.e., the fee minus the rebate) is $0.0001. However, on the taker-maker exchange, the taker gets the rebate of $0.0029 and the liquidity provider pays the fee of $0.0030, whereas on the maker-taker exchange the liquidity provider gets the rebate and the taker pays the fee. This

---

[22] 2015 was the most-recently available full year of data when we began presenting early versions of this research publicly. 2015 is also the best year in terms of data availability for the analysis of exchange speed-technology revenues as described in Appendix B.1.

[23] We have also conducted robustness tests in which we look at the top 1000 symbols by share volume that satisfy these filters except for the $10 million average daily volume filter, which constitutes roughly three-quarters of total volume. Results are qualitatively similar but with more noise.

Figure F.1: Multiple Exchanges at the Same Best Price (Top 5 Exchanges)
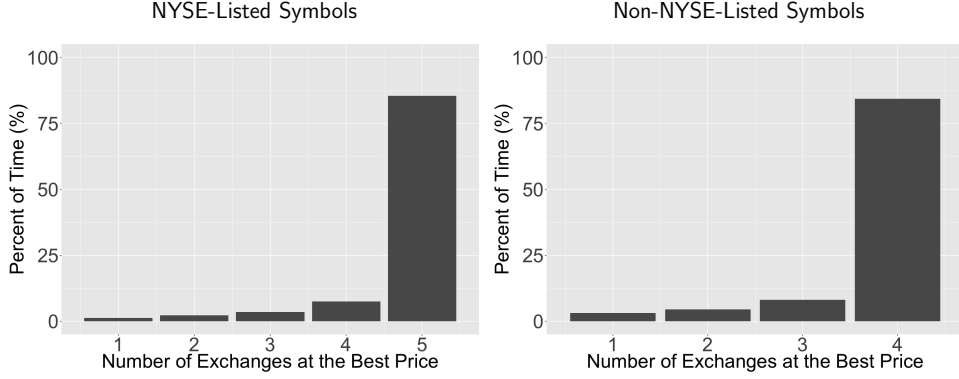
Notes: The data is from NYSE TAQ. Percent of time indicates the percent of symbol-side-milliseconds (e.g. SPY-Bid-10:00:00.001) for which the number of exchanges at the best bid or offer was equal to N. An exchange was at the best price for a symbol-side-millisecond if the best displayed quote on that exchange was equal to the best displayed quote on any of the Top 5 exchanges, all measured at the end of the millisecond. The best bid or offer on the Top 5 exchanges was also the best bid or offer across the Top 8 exchanges in over 99.9% of milliseconds; see Appendix F.1 for details. Sample is 100 highest volume symbols that satisfy data-cleaning filters (see text for description) on all dates in 2015.

means that if the taker transacts on the maker-taker exchange, they in effect pay \$10.00 + \$0.0030 = \$10.0030, whereas if the taker transacts on the taker-maker exchange, they in effect pay \$10.00 - \$0.0029 = \$9.9971. That is, they save \$0.0059 by transacting on the taker-maker exchange. This example illustrates that depth on a taker-maker exchange, when offered, is economically more attractive than depth on a maker-taker exchange, and will be consumed first by optimizing market participants. This effect is observable below.

## F.1 Many Exchanges Simultaneously at the Best Bid and Best Offer

For each symbol i, exchange j, millisecond k, and date t, we compute the exchange's best bid and best offer (ask), denoted BB_{ijkt} and BO_{ijkt}. In case there are multiple quote updates in the symbol-exchange-millisecond, we use the last one. We then compute, for each symbol-millisecond-date, the number of exchanges at the overall best bid and best offer, i.e., we compute:

$$N_{ikt}^{bid} = \sum_j 1\{BB_{ijkt} = \max_{j' \in J} BB_{ij'kt}\} \qquad \text{and} \qquad N_{ikt}^{offer} = \sum_j 1\{BO_{ijkt} = \min_{j' \in J} BO_{ij'kt}\},$$

where J is the set of all exchanges.

As one might expect, the distributions of N_{ikt}^{bid} and N_{ikt}^{offer} are virtually identical, so we combine the data into a single distribution and present it as Figure F.1. We present the results separately for NYSE-listed symbols and non-NYSE listed symbols. The reason for this difference is that, at the time of our data, non-NYSE listed symbols did not trade on NYSE (but did trade everywhere else), whereas NYSE listed symbols traded everywhere. Hence, for NYSE listed symbols the maximum number of exchanges out of the Top 5 that could be at the best bid or offer is 5, whereas for non-NYSE listed symbols the maximum is 4. As can be seen, the modal answer to the question "how many exchanges are at the best price?" is "all of them." For NYSE-listed symbols,

Figure F.2: Multiple Exchanges at the Same Best Price (Top 8 Exchanges)
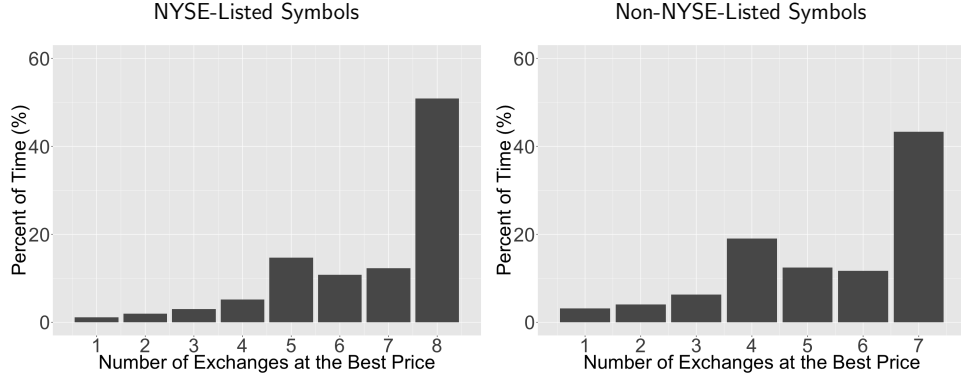
Notes: The data is from NYSE TAQ. Percent of time indicates the percent of symbol-side-milliseconds (e.g. SPY-Bid-10:00:00.001) for which the number of exchanges at the best price was equal to N. The figure considers the Top 8 exchanges; for discussion of Top 8 see the text. An exchange was at the best price for a symbol-side-millisecond if the best displayed quote on that exchange was equal to the best displayed quote on any of the eight exchanges, all measured at the end of the millisecond. Sample is 100 highest volume symbols that satisfy data-cleaning filters (see text for description) on all dates in 2015.

all Top 5 exchanges are at the best bid (similarly, best offer) in 86.1% of milliseconds, and for non-NYSE symbols all 4 exchanges are at the best bid or offer in 84.6% of milliseconds.

Additionally, we also present results for the Top 8 exchanges, i.e., all exchanges with meaningful market share, in Figure F.2. Just as above, we present the results separately for NYSE-listed symbols and non-NYSE listed symbol. For NYSE listed symbols, the maximum number of exchanges out of the Top 8 that could be at the best bid or offer is 8, whereas for non-NYSE listed symbols the maximum is 7.

If we look at the Top 8, all exchanges are at the best bid (similarly, best offer) in about 50% of milliseconds. There is a small peak at 5 exchanges for NYSE-listed stocks and 4 exchanges for non-NYSE-listed stocks. As mentioned above, if a liquidity provider quotes the same price on a taker-maker exchange as on a maker-taker exchange, it is in effect offering a price that is roughly half a penny (the sum of the fee and rebate) better for the taker of liquidity and worse for itself as the provider of liquidity (see Table A.1 in Appendix A.2 for exact numbers). Therefore it makes economic sense that it will often be the case that the best price is found on all of the maker-taker exchanges (5 exchanges for NYSE-listed stocks and 4 exchanges for non-NYSE-listed stocks) but not on the taker-maker exchanges. It is rare that only one or a few exchanges are at the best price.

F.2 Depth Equals Volume

For each symbol i, exchange j, and date t, we compute the exchange's "depth share" and "volume share" for regular-hours trading in that symbol on that date. Volume share, $VolumeShare_{ijt}$, is calculated as the regular-hours volume in shares for symbol i on exchange j on date t divided by total regular-hours volume in symbol i on date t. We calculate depth share, $DepthShare_{ijt}$, by first computing depth for symbol i on exchange j at each millisecond k within the regular-hours trading period of day t, defined as

$$Depth_{ijtk} = \frac{q_{ijtk}^{bid} \cdot 1\{BB_{ijtk} = \max_{j' \in J} BB_{ij'kt}\} + q_{ijtk}^{offer} \cdot 1\{BO_{ijtk} = \min_{j' \in J} BO_{ij'kt}\}}{2},$$

Figure F.3: 2015 Daily Volume Share vs. Depth Share (Top 5 Exchanges)
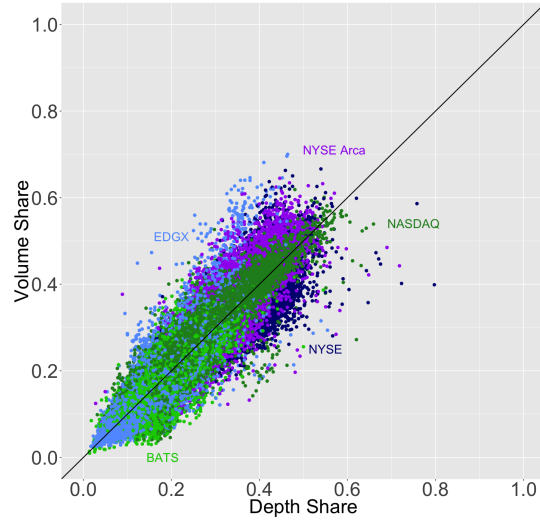
Notes: The data is from NYSE TAQ. The dark line depicts the 45-degree line which is the depth share to volume share relationship predicted by the theory. The results are presented for the Top 5 maker-taker exchanges, and includes the 100 highest volume symbols that satisfy data-cleaning filters on all dates in 2015. Observations are symbol-date-exchange shares, with shares calculated among the Top 5 exchanges.

where q_{ijtk}^{bid} and q_{ijtk}^{offer} denote the quantity at exchange j's best bid and offer for symbol i at millisecond k, and the indicator function requires that j's best bid or offer equals the national best at that millisecond. We then compute the average depth for each symbol-exchange-date by averaging $Depth_{ijtk}$ over all milliseconds, then calculate $DepthShare_{ijt}$ as this average depth divided by the sum of the average depth for each symbol-exchange-date across exchanges. Figure F.3 presents a scatterplot of $VolumeShare_{ijt}$ against $DepthShare_{ijt}$ for the Top 5 exchanges, wherein each dot represents a symbol-exchange-date tuple. We color code by exchange and label each exchange's cluster of dots.
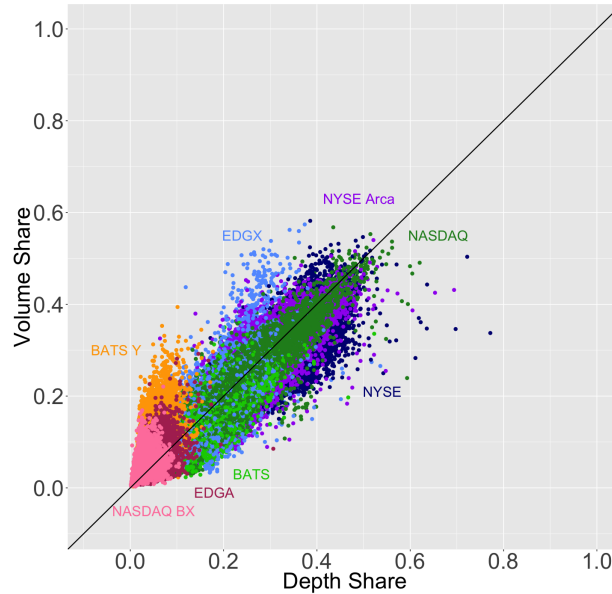
The figure shows that the depth-volume data falls along the 45 degree line for the Top 5 exchanges. The slope of a regression of volume share on depth share is 0.991 (s.e. 0.020), and the R^2 of the relationship is 0.865.[24] In robustness tests, we found that the depth-volume relationship along the 45 degree line obtains at significantly higher frequencies than a day, such as 5 minutes (albeit with more noise), but that at frequencies such as 1 second or 1 millisecond the relationship is not meaningful.[25] This stems from the fact that, at the level of an individual trade, exchange volume shares are often 0% or 100%, so the depth-volume relationship is only meaningful with some aggregation.[26]

---

[24]As a robustness test we looked at the depth-volume relationship for each symbol in our data, running 100 regressions of daily exchange market shares on daily exchange depth shares, one for each symbol. The regression coefficients are very close to one (mean 0.991, st. dev. 0.026) and the R^2 of the relationship is high (mean 0.840, st. dev. 0.136), suggesting that the depth-volume relationship holds at the individual symbol level.

[25]The R^2 of the regression of volume share on depth share is 0.531 at 5 minutes, 0.635 at 10 minutes, 0.745 at 30 minutes, and 0.788 at 1 hour. The regression coefficients are 0.951, 0.957, 0.963 and 0.965 (each statistically indistinguishable from 1).

[26]Our model assumes all investors demand exactly "1" unit of perfectly-divisible liquidity and in equilibrium exactly 1 unit of liquidity is offered across exchanges so investors must spread their demand across multiple exchanges. In reality, investors demand varying amounts of liquidity. Investors who only wish to trade a small amount (e.g., 100 shares) often do so with a single small trade on a single exchange. Investors who wish to trade a larger amount often break their total

Figure F.4: 2015 Daily Volume Share vs. Depth Share: Top 8 Exchanges

Notes: The data is from NYSE TAQ for the Top 8 exchanges. The dark line depicts the 45-degree line which is the depth share to volume share relationship predicted by the theory. Observations are symbol-date-exchange shares, with shares calculated among the Top 8. For details of share calculations see the text. Sample is 100 highest volume symbols that satisfy data-cleaning filters (see text for description) on all dates in 2015.

We also present a scatterplot of $VolumeShare_{ijt}$ against $DepthShare_{ijt}$ for the Top 8 exchanges in Figure F.4. The figure shows that most of the depth-volume data falls along the 45 degree line. Formally, the slope of a regression of volume share on depth share is 0.991 (s.e. 0.020), and the R^2 of the relationship is 0.865. Just as in Figure F.3 above, the Top 5 exchanges – NYSE, Nasdaq, NYSE Arca, BZX (sometimes referred to simply as BATS), and EDGX – are tightly scattered along the 45 degree line (note that shares for each day are recomputed with the addition of the 3 taker-maker exchanges). The taker-maker exchanges – EDGA, BYX (sometimes referred to as BATS Y), Nasdaq BX – are clustered at the bottom left of the figure and have a steeper slope than the maker-taker exchanges. That is, the taker-maker exchanges have volume shares that are typically greater than depth shares. The reason for this, as mentioned in the example above, is that the taker-maker exchanges pay a rebate to the taker of liquidity. Thus, while depth on taker-maker exchanges is comparatively rare, when there is depth on taker-maker exchanges it is more economically attractive, after accounting for fees, than depth at the same pre-fee price on the larger maker-taker exchanges. For this reason, it makes sense that taker-maker exchanges have volume shares that are larger than their depth shares.

---

desired quantity into smaller individual orders spread out over time. So, volume shares at the trade-by-trade level are often 100% for a single exchange and 0% for all others. In addition, there will be some "excess supply" in the data purely due to tick sizes. Without tick sizes, if the smallest two investors demand 100 and 200 shares, the first 100 shares will be split across exchanges and the second 200 shares will be split across exchanges at a slightly wider spread. However with tick sizes, both may end up quoted at the same price and an investor that demands 200 shares may be able to buy all 200 on a single exchange. However, the logic of our model suggests that in both cases, at a higher level of aggregation, volume shares should match depth shares — else, the marginal unit of liquidity will be too adversely selected on some exchanges and too favorable on others.

G   Discussion of Discrete vs. Continuous with Tick-Size Constraints and Agency Frictions

In this appendix we discuss competition between a Discrete exchange and one or more Continuous exchanges when there are tick-size constraints and agency frictions. The purpose of this appendix, as mentioned in Section 4.2, is to provide a back-of-envelope calculation for an appropriate exclusivity period while taking into account real-world frictions.

For this discussion we make the following assumptions:

- Tick-size constraints. We assume that stocks trade in increments of $ticksize = \$0.01$. This reflects tick-size regulations in U.S. stock markets for stocks with a nominal share price greater than $1 (Reg NMS Rule 612).

- Agency frictions. We assume that whenever there is liquidity at the same quoted price on Continuous and Discrete, investors always break ties in favor of Continuous. This tie-breaking in favor of Continuous is independent of any fee differences (discussed below). Investors route orders to Discrete only if Discrete has a quoted price that is at least a full tick better for the investor, i.e., when they are mandated to do so under the Reg NMS order protection rule. This assumption is meant to capture, in a simple and worst-case way, any agency frictions that might favor trading on Continuous markets over Discrete markets (in the spirit of Battalio, Corwin and Jennings, 2016).

- Maker-Taker fees. We assume that the Continuous market uses a maker-taker fee schedule with a take fee equal to the regulatory maximum under the Reg NMS Access Rule of +30 mills ($0.0030) per share, and a make fee equal to -30 mills per share (i.e., make rebate of 30 mills), for a net fee of 0. This approximates current practice as discussed in Appendix A.2 and Appendix F. Additionally, in conjunction with the tie-breaking assumption in the previous bullet point, this assumption about fees is a worst-case for Discrete. Since we have assumed that investors only route to Discrete when liquidity is a full tick more attractively priced, Discrete would like to charge investors a higher fee than they pay on Continuous — investors would be willing to pay a higher fee conditional on trading, since whenever they trade on Discrete they are saving a full tick. However, the Continuous market is already charging investors the regulatory maximum. Therefore, if Discrete is to charge a positive fee, it does so by charging a take fee of 30 mills, just as on Continuous, and a make fee of $-(30 - f_D)$ mills (i.e., rebate of $30 - f_D$ mills), where f_D denotes the net fee to Discrete per share.

- Distribution of fundamental values. For the purpose of discussion, we assume that the fundamental value y of the security is uniformly distributed such that all values between any two relevant ticks are equally likely.

- Magnitude of latency arbitrage. For the purpose of discussion, we utilize the estimate from Aquilina, Budish and O'Neill (2022) that the latency arbitrage tax on liquidity is 0.42 basis points (0.0042%) of traded volume.[27]

---

[27] As shown in Section 5.5 of Aquilina, Budish and O'Neill (2022), for the purpose of computing the amount by which eliminating latency arbitrage would reduce the cost of liquidity, it is technically more accurate to use latency-arbitrage profits as a proportion of non-race traded volume as opposed to as a proportion of all traded volume. This non-race volume latency arbitrage tax is 0.53 basis points as opposed to 0.42 basis points. To err on the side of conservatism we use the 0.42 basis points figure.

Given our assumptions about tick-sizes and fees, if there were a single Discrete exchange operating in isolation, the equilibrium best offer would be $\left\lceil y + \frac{s^*_{discrete}}{2} - (\$0.0030 - f_D) \right\rceil$, where the notation $\lceil x \rceil$ denotes rounding up to the nearest whole-penny increment. (For simplicity, we focus discussion on just the offer, the bid being symmetric.) If there were a single Continuous exchange operating in isolation, the equilibrium best offer would be $\left\lceil y + \frac{s^*_{continuous}}{2} - (\$0.0030) \right\rceil$. TFs would only be able to offer liquidity on Discrete at a strictly better whole-penny increment than on continuous if

$$\left\lceil y + \frac{s^*_{discrete}}{2} - (\$0.0030 - f_D) \right\rceil < \left\lceil y + \frac{s^*_{continuous}}{2} - (\$0.0030) \right\rceil, \tag{G.1}$$

that is, if the magnitude of latency arbitrage, represented by $\frac{s^*_{continuous}}{2} - \frac{s^*_{discrete}}{2}$, is large enough to "cross a tick" given the current fundamental value y, and accounting for any difference in fees and rebates.

Given the assumption that fundamental values are uniformly distributed between any two relevant ticks, the probability that the fundamental value y satisfies condition (G.1) is:

$$\frac{\frac{s^*_{continuous}}{2} - \frac{s^*_{discrete}}{2} - f_D}{ticksize}, \tag{G.2}$$

truncated below at 0 (in case f_D is too large) and above at 1 (in case $\frac{s^*_{continuous}}{2} - \frac{s^*_{discrete}}{2} - f_D$ exceeds the tick size). For example, if $\frac{s^*_{continuous}}{2} - \frac{s^*_{discrete}}{2} - f_D = \$0.0020$, then the probability that y is such that condition (G.1) holds is 20%. Our assumption about the magnitude of latency arbitrage enables us to compute Discrete's market share for symbol i from (G.2) as:

$$FBAshare_i = \frac{(0.0042\% \cdot shareprice_i - f_D)}{\$0.01} \tag{G.3}$$

again truncating below at 0 and above at 1.

We use TAQ data in 2015 to compute (G.3) for all symbols with $shareprice_i > \$5$ that traded continuously throughout the year under the same ticker, with $shareprice_i$ calculated as the volume-weighted average trade price for symbol i and net fees per share f_D ranging from 0 mills to 30 mills. A net fee f_D of 0 mills corresponds to a take fee of +30 mills and a make fee of -30 mills (i.e., a rebate) as on the Continuous market, whereas a net fee f_D of 30 mills corresponds to a take fee of +30 mills and a make fee of 0. We then use $FBAshare_i$ to compute overall FBA market shares by share volume and dollar volume (relative to the symbol universe). We also use $FBAshare_i$ to compute annual FBA revenue, which is $FBAshare_i$ multiplied by f_D and by overall share volume for symbol i, then summed over all symbols. The results are summarized in Table G.1.

At a net fee of $f_D = 0$, i.e., the same maker-taker fee structure as the Continuous market, the Discrete exchange's share is computed as 18.7% in share volume and 37.2% in dollar volume. The reason why dollar volume is meaningfully higher than share volume is that tick-size constraints are less binding for high nominal share price stocks.

At a net fee of $f_D = 10$ mills, the Discrete exchange's share is computed as 10.4% in share volume and 27.9% in dollar volume. Fee revenues are \$92.1 million per year. Fee revenues are maximized at about \$106 million per year, using a net fee of about 20 mills.

Clearly, this exercise is very back-of-the-envelope. In particular, it utilizes a magnitude for latency arbitrage taken from UK equity markets, whereas if better data were available in U.S. equity markets it would be possible to directly calculate both the average level and the cross-sectional heterogeneity across stocks and ETFs

Table G.1: FBA Market Share and Revenue with Tick-Size Constraints and Agency Frictions

FBA Net Fee Per Share (Mills) (*)	FBA Share (% of Share Volume)	FBA Share (% of Dollar Volume)	FBA Annual Revenue ($ Millions)
0	18.7%	37.2%	0.0
1	17.7%	36.2%	15.7
2	16.7%	35.2%	29.6
3	15.8%	34.3%	41.8
4	14.8%	33.3%	52.5
5	14.0%	32.4%	61.8
6	13.2%	31.4%	69.9
7	12.4%	30.5%	76.8
8	11.7%	29.6%	82.8
9	11.0%	28.8%	87.8
10	10.4%	27.9%	92.1
15	7.8%	23.9%	103.8
20	6.0%	20.5%	105.9
25	4.7%	17.6%	103.0
30	3.7%	15.2%	97.2

(*) FBA Net Fee = Take Fee + Make Fee. As discussed in the text, we hold fixed the Take Fee at 30 mills ($0.0030) per share, and vary the Make Fee from -30 mills (i.e., rebate of $0.0030 per share) to 0.

Notes: Data from NYSE TAQ in 2015. The symbol universe is all stocks with $shareprice_i > \$5$ that traded continuously throughout the year under the same ticker. We first calculate $FBAshare_i$ for each symbol i according to equation (G.3), with $shareprice_i$ calculated as the volume-weighted average trade price of the symbol over all trading days in 2015, and f_D set to the values in the "FBA Net Fee Per Share" column. We then use the FBA market shares for each symbol to compute overall FBA market shares by share volume and dollar volume, with overall market shares expressed relative to the symbol universe. FBA annual revenue is computed as overall FBA share volume times the net fee f_D.

of latency arbitrage in the U.S.. It also incorporates complicated market frictions in a very stylized manner. Nevertheless, we hope that this exercise provides the reader with a useful sense of magnitudes both for interpreting the Discrete vs. Continuous theoretical results in Section 3.6 and for thinking about the market design exclusivity period idea in Section 4.2.

H Institutional Background: Market Design Proposals that Relate to Latency Arbitrage

To date, there have been four exchange design proposals in the United States stock market that relate to latency arbitrage: the Investors' Exchange (IEX) symmetric speed bump proposal in 2016, the Chicago Stock Exchange (CHX) asymmetric delay proposal in 2016, the Cboe EDGA asymmetric delay proposal in 2019, and the Cboe BYX user-initiated batch auction proposal in 2020. This level of market design innovation activity is consistent with latency arbitrage being important in practice. That said, a salient feature of these proposals is that only CHX's would have protected displayed quotes from latency arbitrage — that is, only CHX's proposal concerned the type of quotes that are at the heart of price discovery in the U.S. stock market, and this from a tiny venue with <1% market share at the time. After some regulatory back-and-forth and an official "staying" of the proposal, CHX was acquired by the New York Stock Exchange Group in 2017, who then withdrew the proposal. The other three proposals all relate to latency arbitrage but for non-displayed and/or non-protected quotes, which in a loose economic sense means the quotes free ride off of the price discovery from displayed, protected quotes on the same or other venues (see Antill and Duffie (2018) and Zhu (2014) for formal models that relate to this informal idea, and see Budish (2016b) for a regulatory comment letter that relates to this idea). These proposals seem to capture the spirit of the "puppy dog ploy" in Fudenberg and Tirole (1984) or the "judo economics" of Gelman and Salop (1983) where a smaller exchange can win market share but not enough to induce a full competitive response from larger incumbent exchanges.

This appendix provides details and references for each of these four exchange design proposals. Throughout, we use primary sources where possible, though in some instances we fill in details based on contemporaneous discussions with industry participants about the relevant proposals. The appendix also discusses a recent innovation of IEX in the form of a new order type called D-Limit. This order type does protect displayed, Reg NMS protected orders from some kinds of latency arbitrage, namely those latency arbitrage opportunities IEX's statistical models detect based on price updates from other exchanges.

CHX Asymmetric Delay. In 2016, the Chicago Stock Exchange (CHX) proposed to adopt a 350 microsecond asymmetric speed bump for its exchange. As modeled formally in Baldauf and Mollner (2020) and discussed informally in Budish, Cramton and Shim (2015), an asymmetric speed bump (or "asymmetric delay"), in which liquidity taking orders are delayed but liquidity providing orders and cancelations are not, protects displayed orders against latency arbitrage. Intuitively, the asymmetric delay gives liquidity providers canceling stale quotes a tiny head start in a sniping race. See Budish (2016a) for details and discussion of CHX's proposal.

The proposal went through two iterations, the first called "Liquidity Taking Access Delay" (LTAD) and the second called "Liquidity Enhancing Access Delay" (LEAD). The proposals were substantively similar in the design of the speed bump, the difference being that the second proposal attempted to provide the protection only for participants who were designated liquidity providers on CHX (with some formal obligations associated with that designation), based on feedback the exchange got to its first proposal from the SEC.

Both proposals were met with fierce resistance from the larger exchanges and several high-frequency trading firms. For example, Citadel wrote that it "unfairly structurally and systematically discriminates against market participants that are primarily liquidity takers, such as retail investors" (Citadel, 2016). See U.S. Securities and Exchange Commission (2017a, 2018a) for the full comment files. The main substantive argument against the proposal expressed in public comment letters was that the *asymmetry* of the delay is inconsistent with the

fairness provisions of the 1934 Exchange Act.

CHX's second proposal was officially "stayed" by the SEC in Oct 2017, which is a regulatory maneuver that does not officially rule on a proposal but does signal a level of opposition to it. CHX was then acquired by the New York Stock Exchange Group, and officially withdrew the proposal in July 2018 (CHX, 2018; Michaels and Osipovich, 2018).

IEX Symmetric Speed Bump. The Investors' Exchange (IEX), famously profiled in Michael Lewis's *Flash Boys*, proposed to become a new exchange in 2015 with a market design that incorporates a 350 microsecond symmetric speed bump (see IEX's initial Form 1 filing at IEX (2015a) and subsequent amendments at IEX (2016)). Importantly, a symmetric speed bump, which delays all incoming messages equally, has no effect on latency-arbitrage races in the displayed, continuous limit order book market. As one high-frequency trading firm analogized in a comment letter critical of IEX, it is like taking the 100-meter dash in the olympics, and adding 0.000350 seconds to all times — it's still a race, with the same winner, only it takes a bit longer (Hudson River Trading, 2015).

The value of IEX's symmetric speed bump is the protection from latency arbitrage it created for non-displayed, pegged orders. Such orders are re-priced based on the best bid or offer on exchanges other than IEX — and, crucially, IEX receives those updates in less time than the 350 microseconds of delay (roughly 200 microseconds, based on the geography of exchange data centers in New Jersey). Thus, if there is public information that affects prices on other exchanges, IEX can re-price pegged orders on its exchange faster than a sniper could trade against those orders in response to the public information. Non-displayed orders currently are the source of the majority of IEX's trading volume.[28]

IEX's exchange application was controversial, eliciting an unprecedented 477 SEC comment letters, many of which were overtly hostile (U.S. Securities and Exchange Commission, 2016a). One particularly memorable letter was from the New York Stock Exchange, which wrote, "Like the non-fat yogurt shop on Seinfeld, which actually serves tastier, full-fat yogurt to increase its sales, IEX advertises that it is A Fair, Simple, Transparent Market whereas it proposes rules that would make IEX an unfair, complex, and opaque exchange" (NYSE, 2015b). Our sense is that this hostility to IEX reflected a genuine mixture of incumbents' desire to preserve the status quo and legitimate concerns about the details of IEX's market design. The SEC approved IEX's application in June 2016, while also issuing a rules clarification about the use of de minimis delays in exchange designs (see discussion above in Appendix D.2).

See Budish (2016b) for further details and discussion of IEX's market design.

Cboe EDGA Asymmetric Speed Bump In 2019, Cboe's EDGA exchange proposed an asymmetric speed bump. Unlike CHX's proposal, EDGA proposed for the exchange *not* to have Reg NMS quote protection, and proposed a delay amount, 4 milliseconds, that was longer than the de minimis delay threshold of 1 millisecond adopted by the SEC in June 2016. The rationale for the 4 millisecond amount was "The Exchange believes that this delay would negate the advantages that opportunistic trading firms that use the latest microwave connections [from Chicago to New Jersey] have over liquidity providers using traditional fiber connections." (4 milliseconds is roughly the difference in one-way travel time between a microwave link from Chicago to New Jersey and a traditional fiber-optic cable connection).

---

[28]For example, during the period Jan-June 2020, non-displayed orders were over 85% of IEX's volume. In 2021, this percentage has declined to about 70% of volume, likely because of the D-Limit displayed order type discussed below. See https://iextrading.com/stats/ for the underlying data.

EDGA's proposal was rejected by the SEC in Feb 2020 on the grounds that the delay is "discriminatory" (against executable orders relative to liquidity providing orders) without adequate justification. In particular, the SEC wrote that "the Commission does not believe that the Exchange has supported its assertions and demonstrated that the [proposed] delay mechanism is appropriately tailored to address latency arbitrage and not permit unfair discrimination" (U.S. Securities and Exchange Commission, 2020c).

It remains an open question whether the SEC is open to an asymmetric delay mechanism with a shorter delay window and more detailed empirical justification than EDGA provided, or whether the asymmetry is viewed as intrinsically unfair and hence inconsistent with the fairness language in the 1934 Exchange Act (in contrast to the symmetric speed bump adopted by IEX, which was approved).

Cboe BYX Batch Auctions   In 2020, Cboe BYX proposed to incorporate user-initiated 100 millisecond batch auctions alongside its continuous limit order book. The proposal was approved by the SEC and as of July 2021 is scheduled to go live in Q4 2021.[29]

The auction design details have several similarities with the frequent batch auctions proposal in Budish, Cramton and Shim (2015): the auction per se is a standard uniform-price double auction; orders are simple price-quantity pairs, can be modified or canceled at any time, and remain open until either executed or canceled; the auction duration is 100 milliseconds, similar to the time intervals discussed in Budish, Cramton and Shim (2015).

There are two key differences. First, the BYX auction is designed to run alongside the BYX continuous market, with the auctions only run when a user initiates one by submitting an auction-only order that crosses the market (e.g., against another auction-only order resting at the midpoint of the continuous market's best bid and offer). In contrast, in the Budish, Cramton and Shim (2015) frequent batch auction design, the auctions run continuously throughout the day — e.g., if the batch interval is 1 millisecond, there are 23.4 million frequent batch auctions run throughout the day — in place of the continuous limit order book.

The second, and central, difference is information policy. In the BYX design, auction orders are all *hidden*. The only displayed orders on BYX are in the traditional, continuous limit order book market. If a user initiates an auction, by submitting a new auction-only order that crosses against either a displayed limit order (though, here, the user could just immediately trade against that displayed limit order if they wanted to, in the continuous market) or against a previously placed, hidden, auction-only order (e.g., at the midpoint), then BYX sends out a public message that an auction has been initiated in the relevant symbol. Participants then have a random amount of time, of less than 100 milliseconds, to respond with any additional orders they wish to add to the auction. During this time, BYX disseminates indicative price and quantity information every millisecond.

Substantively, what this information policy means is that it is not possible to provide displayed liquidity on the BYX auction that is protected from latency arbitrage. The only displayed liquidity in the BYX market is the traditional displayed limit orders, on the continuous part of the market, and these are vulnerable to sniping just like other exchanges' displayed limit orders.

In contrast, in the Budish, Cramton and Shim (2015) frequent batch auction design, the information policy is that all orders resting in the frequent batch auction market are displayed publicly after every discrete time interval. So, it is possible to provide liquidity that is displayed to the market, while being protected from latency

---

[29]See U.S. Securities and Exchange Commission (2020b) for the formal rules filing, Cboe Global Markets, Inc. (2021b) for a plain-English description of the auction design, and U.S. Securities and Exchange Commission (2021) for the SEC approval. The announcement that the auction is scheduled to go live in Q4 2021 is the July 19th post at Cboe Global Markets, Inc. (2021a).

arbitrage in case there is public information that triggers a sniping race.[30]

Another potential issue with the BYX information policy is that it could create a form of information leakage. Specifically, in the 100 milliseconds before the auction executes, there is public information disseminated to the market about indicative prices and quantities. So, for example, if a large quantity is forecast to execute on BYX, participants could trade in advance of this on other venues. Or, a participant could create the impression that a large quantity will execute on the BYX auction, but then immediately cancel the auction order. BYX appears to be aware of these issues and attempts to prohibit these kinds of behaviors.[31] Again, in Budish, Cramton and Shim (2015)'s proposal, information is disseminated in discrete time, so if there are new orders that will cause a large quantity to execute in the auction, the market only learns of this after the auction has actually executed. And, if a new order is entered that would cause a large quantity to execute, but is then canceled before the auction interval ends, the market is not misleadingly alerted that a large trade will occur.

The BYX proposal is similar to batch auction designs that have been implemented in European equity markets. Such batch auction venues currently have market share of several percentage points in European equities, primarily facilitating matching at the midpoint of the displayed market's bid and offer. See the European Securities and Markets Authority (2019) for some additional details and see Budish (2019b) for a discussion of the strengths and weaknesses of these auction designs.

IEX D-Limit Order Type   IEX recently received regulatory approval for a new order type on its exchange, called D-Limit, which protects displayed, Reg NMS protected limit orders from latency arbitrage that IEX is able to detect using a published mathematical formula (U.S. Securities and Exchange Commission, 2019b). Specifically, IEX uses an indicator called "Crumbling Quote Indicator" (CQI), based on prices from other exchanges, and if the indicator signals that prices are moving adversely on other exchanges, IEX adjusts the D-Limit order to be priced less aggressively by one tick. IEX provides some execution statistics that suggest that D-Limit orders experience less adverse selection (less price impact) than traditional limit orders, while also being less likely to execute at their original price, both of which are consistent with the order type helping avoid some forms of sniping (Ryan, 2021). IEX also reports that the share of its volume that is displayed as opposed to non-displayed has increased from about 10% to about 30%, which also seems consistent with the idea that with IEX is now able to protect both non-displayed orders and displayed orders from some forms of sniping (Basar, 2021).

As a solution to latency arbitrage, the D-Limit order type has two related weaknesses relative to a change in market design per se. First, it is based on a hard-coded statistical formula (roughly speaking, think of it as a linear combination of price updates from BATS, NYSE, and Nasdaq). Second, and more centrally, its approach to reducing latency arbitrage explicitly relies on price updates from other exchanges, that is, free-riding off of

---

[30]It is important to keep in mind that in most symbols, in most fractions of a second, not much happens; e.g., if the batch interval is 1 millisecond, then in most of the 23.4 million auction intervals per symbol, the market is simply disseminating the state of the auction's order book, without any trades or book updates occurring. If a participant wants to "buy 100 shares at the offer", they can do so, in essentially the same way as in the continuous limit order book market. The difference, again, is how this market design processes bursts of activity in an interval, in case there is public information that triggers a race.

[31]From pg. 21-22 of the formal rules filing: "A pattern or practice of submitting orders for the purpose of disrupting or manipulating Periodic Auctions, including entering and immediately cancelling Periodic Auction Orders, would be deemed conduct inconsistent with just and equitable principles of trade. The Exchange would conduct surveillance to ensure that Users do not inappropriately enter Periodic Auction Orders for impermissible purposes, such as to gain information about other Periodic Auction Orders that are resting on the Periodic Auction Book, or otherwise disrupting or manipulating Periodic Auctions." (See U.S. Securities and Exchange Commission (2020b).)

As Roth and Xing (1997) wrote, "rules are data".

price discovery elsewhere. As a thought experiment, imagine if IEX had 100% market share for a given symbol. In this case, its market design would not offer any protection from sniping (e.g., from signals generated by correlated assets), because there would not be any prices elsewhere to free-ride off of.[32]

That said, the D-Limit order type represents the first time there have been displayed, Reg NMS protected limit orders that are protected from at least some forms of latency arbitrage. This makes it a notable innovation.

---

[32]This is the same limitation Budish (2016b) notes about IEX's approach to protecting non-displayed orders from latency arbitrage, by pegging their price to the BBO on other exchanges.

References

Amihud, Yakov, and Haim Mendelson. 1996. "A New Approach to the Regulation of Trading Across Securities Markets." *New York University Law Review*, 71(6): 1411–1466.

Antill, Samuel, and Darrell Duffie. 2018. "Augmenting Markets with Mechanisms." NBER Working Paper No. 24146.

Aquilina, Matteo, Eric Budish, and Peter O'Neill. 2022. "Quantifying the High-Frequency Trading 'Arms Race'." *The Quarterly Journal of Economics*, 137(1): 493–564.

Baldauf, Markus, and Joshua Mollner. 2020. "High-Frequency Trading and Market Performance." *The Journal of Finance*, 75(3): 1495–1526.

Basar, Shanny. 2021. "IEX Increases Lit Volume With D-Limit Order." February 16, Retrieved July 27, 2021 from `https://www.marketsmedia.com/iex-increases-lit-volume-with-d-limit-order/`.

BATS Global Markets, Inc. 2015*a*. "BATS BYX Exchange Fee Schedule." Retrieved April 11, 2017 from `http://www.bats.com/us/equities/membership/fee_schedule/byx/` through Wayback Machine (archived on February 28, 2015).

BATS Global Markets, Inc. 2015*b*. "BATS BZX Exchange Fee Schedule." Retrieved April 12, 2017 from `http://www.bats.com/us/equities/membership/fee_schedule/bzx/` through Wayback Machine (archived on February 28, 2015).

BATS Global Markets, Inc. 2015*c*. "EDGA Exchange Fee Schedule." Retrieved April 11, 2017 from `http://www.bats.com/us/equities/membership/fee_schedule/edga/` through Wayback Machine (archived on February 28, 2015).

BATS Global Markets, Inc. 2015*d*. "EDGX Exchange Fee Schedule." Retrieved April 11, 2017 from `http://www.bats.com/us/equities/membership/fee_schedule/edgx/` through Wayback Machine (archived on April 27, 2015).

Battalio, Robert, Shane A. Corwin, and Robert Jennings. 2016. "Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality." *Journal of Finance*, 71(5): 2193–2238.

Budish, Eric. 2016*a*. "Re: Chicago Stock Exchange Liquidity Taking Access Delay (Release No. 34-78860; SR-CHX-2016-16)." Retrieved February 12, 2019 from `https://www.sec.gov/comments/sr-chx-2016-16/chx201616-9.pdf`.

Budish, Eric. 2016*b*. "Re: Investors' Exchange LLC Form 1 Application (Release No. 34-75925; File No. 10-222)." Retrieved December 22, 2018 from `https://www.sec.gov/comments/10-222/10222-371.pdf`.

Budish, Eric. 2016*c*. "Re: Proposed Commission Interpretation Regarding Automated Quotations Under Regulation NMS (Release No. 34-77407; File No. S7-03-16)."

Budish, Eric. 2019*b*. "Response to ESMA's Call for Evidence: "Periodic Auctions for Equity Instruments" (ESMA70-156-785)." Retrieved June 9, 2021 from `https://faculty.chicagobooth.edu/eric.budish/research/Budish-comment-ESMA-periodic-auctions-Jan2019.pdf`.

Budish, Eric, Peter Cramton, and John Shim. 2015. "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." *The Quarterly Journal of Economics*, 130(4): 1547–1621.

Canadian Securities Administrators. 2009. "Notice of Amendments to National Instrument 21-101 Marketplace Operation and National Instrument 23-101 Trading Rules." *Ontario Securities Commission*, Retrieved April 17, 2019 from `http://www.osc.gov.on.ca/documents/en/Securities-Category2/rule_20091113_21-101_new-noa-21-101and23-101.pdf`.

Cboe Global Markets, Inc. 2021*a*. "Cboe BYX Equities Periodic Auctions Launch Update." July 19, Retrieved July 27, 2021 from `https://cdn.cboe.com/resources/release_notes/2021/Cboe-BYX-Equities-Periodic-Auctions-Launch-Update.pdf`.

Cboe Global Markets, Inc. 2021*b*. "Cboe U.S. Equities Periodic Auctions." Introduction Website, Retrieved July 27, 2021 from `https://www.cboe.com/us/equities/trading/offerings/periodic_auctions/`.

CHX. 2018. "Re: Release No. 34-82034; In the Matter of the Chicago Stock Exchange, Inc. ("Exchange") - For an Order Granting the Approval of Proposed Rule Change to Adopt the CHX Liquidity Enhancing Access Delay ("LEAD") on a Pilot Basis (File No. SR-CHX-2017-04)." Retrieved February 25, 2019 from `https://www.sec.gov/comments/sr-chx-2017-04/chx201704-4118079-171622.pdf`.

Citadel. 2016. "Re: Proposed CHX Liquidity Taking Access Delay (Release No. 34-78860; File No. SRCHX-2016-16)." Retrieved February 15, 2019 from `https://www.sec.gov/comments/sr-chx-2016-16/chx201616-7.pdf`.

European Securities and Markets Authority. 2019. "Final Report: Call for Evidence on Periodic Auctions." June 11, Retrieved July 27, 2021 from `https://www.esma.europa.eu/sites/default/files/library/esma70-156-1035_final_report_call_for_evidence_periodic_auctions.pdf`.

Fudenberg, Drew, and Jean Tirole. 1984. "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look." *American Economic Review: Papers and Proceedings*, 74(2): 361–366.

Gelman, Judith R., and Steven C. Salop. 1983. "Judo Economics: Capacity Limitation and Coupon Competition." *The Bell Journal of Economics*, 14(2): 315–325.

Hasbrouck, Joel, George Sofianos, and Deborah Sosebee. 1993. "New York Stock Exchange Systems and Trading Procedures." NYSE Working Paper No. 93-01.

Hudson River Trading. 2015. "Re: Investors' Exchange LLC Form 1 Application (Release No. 34-75925; File No. 10-222)." December 4. Public Statement. Retrieved July 27, 2021 from `https://www.sec.gov/comments/10-222/10222-33.pdf`.

IEX. 2015*a*. "Investors' Exchange LLC - Form 1 Application and Exhibits, Release No. 34-75925; File No. 10-222." September 16. IEX Application Form 1. Retrieved July 28, 2021 from `https://www.sec.gov/rules/other/2015/investors-exchange-form-1.htm`.

IEX. 2016. "Investors' Exchange LLC - Form 1 Application and Exhibits, Release No. 34-75925; File No. 10-222." March 18. IEX Application Form 1. Retrieved July 28, 2021 from `https://www.sec.gov/rules/other/2016/investors-exchange-form-1-amendments.htm`.

MacKenzie, Donald. 2021. *Trading at the Speed of Light: How Ultrafast Algorithms Are Transforming Financial Markets.* Princeton University Press.

Michaels, Dave, and Alexander Osipovich. 2018. "NYSE in Talks to Buy Chicago Stock Exchange." *Wall Street Journal*, March 30.

Nasdaq, Inc. 2015*a*. "Nasdaq BX Fee Schedule." Retrieved April 11, 2017 from `https://www.nasdaqtrader.com/Trader.aspx?id=bx_pricing` through Wayback Machine (archived on April 1, 2015).

Nasdaq, Inc. 2015*b*. "Price List - Trading Connectivity." Retrieved April 12, 2017 from `http://www.nasdaqtrader.com/Trader.aspx?id=PriceListTrading2` through Wayback Machine (archived on April 8, 2015).

NYSE. 2015*a*. "Price List 2015." Retrieved April 12, 2017 from `https://www.nyse.com/publicdocs/nyse/markets/nyse/NYSE_Price_List.pdf` through Wayback Machine (archived on September 1, 2015).

NYSE. 2015*b*. "Re: Investors' Exchange LLC Form 1 Application (Release No. 34-75925; File No. 10-222)." Retrieved January 24, 2019 from `https://www.sec.gov/comments/10-222/10222-19.pdf`.

NYSE Arca Equities, Inc. 2015. "Schedule of Fees and Charges for Exchange Services." Retrieved April 19, 2017 from `https://www.nyse.com/publicdocs/nyse/markets/nyse-arca/NYSE_Arca_Marketplace_Fees.pdf` through Wayback Machine (archived on August 3, 2015).

O'Hara, Maureen, and Jonathan R. Macey. 1997. "The Law and Economics of Best Execution." *Journal of Financial Intermediation*, 6(3): 188–223.

Petrella, Giovanni. 2010. "MiFID, Reg NMS and Competition across Trading Venues in Europe and the USA." *Journal of Financial Regulation and Compliance*, 18(3): 257–271.

Roth, Alvin, and Xiaolin Xing. 1997. "Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists." *Journal of Political Economy*, 105(2): 284–329.

Ryan, Ronan. 2021. "D-Limit: One of These Things is Not Like the Others." February 1, Retrieved July 27, 2021 from `https://medium.com/boxes-and-lines/d-limit-one-of-these-things-is-not-like-the-others-3c6b438d3a6e`.

U.S. Congress. 1994. "Unlisted Trading Privileges Act of 1994." H.R. 4535. 103rd Congress. Retrieved January 4, 2019 from `https://www.congress.gov/bill/103rd-congress/house-bill/4535/text`.

U.S. Securities and Exchange Commission. 1994. "Market 2000: An Examination of Current Equity Market Developments."

U.S. Securities and Exchange Commission. 2000. "Final Rule: Unlisted Trading Privileges, SEC Release No. 34-43217." Retrieved December 22, 2018 from `https://www.sec.gov/rules/final/34-43217.htm`.

U.S. Securities and Exchange Commission. 2005. "Regulation NMS, SEC Release No. 34-51808." Retrieved November 14, 2018 from `https://www.sec.gov/rules/final/34-51808.pdf`.

U.S. Securities and Exchange Commission. 2016*a*. "Comments on Investors' Exchange LLC; Notice of Filing of Application, as Amended, for Registration as a National Securities Exchange under Section 6 of the Securities Exchange Act of 1934." Retrieved July 28, 2021 from `https://www.sec.gov/comments/10-222/10-222.shtml`.

U.S. Securities and Exchange Commission. 2016*c*. "Staff Guidance on Automated Quotations under Regulation NMS." Retrieved November 4, 2019 from `https://www.sec.gov/divisions/marketreg/automated-quotations-under-regulation-nms.htm`.

U.S. Securities and Exchange Commission. 2017*a*. "Comments on CHX Rulemaking: Notice of Filing of Proposed Rule Change to Adopt the CHX Liquidity Taking Access Delay." Retrieved December 22, 2018 from `https://www.sec.gov/comments/sr-chx-2016-16/chx201616.shtml`.

U.S. Securities and Exchange Commission. 2018*a*. "Comments on CHX Rulemaking: Notice of Filing of Proposed Rule Change to Adopt the CHX Liquidity Enhancing Access Delay." Retrieved December 22, 2018 from `https://www.sec.gov/comments/sr-chx-2017-04/chx201704.htm`.

U.S. Securities and Exchange Commission. 2019*b*. "Notice of Filing of Proposed Rule Change to Add a New Discretionary Limit Order Type." December 20. IEX D Limit Filing. Retrieved July 28, 2021 from `https://www.sec.gov/rules/sro/iex/2019/34-87814.pdf`.

U.S. Securities and Exchange Commission. 2020*b*. "Notice of Filing of a Proposed Rule Change to Introduce Periodic Auctions for the Trading of U.S. Equity Securities." July 29, Retrieved July 28, 2021 from `https://www.sec.gov/rules/sro/cboebyx/2020/34-89424.pdf`.

U.S. Securities and Exchange Commission. 2020*c*. "Order Disapproving Proposed Rule Change to Introduce a Liquidity Provider Protection Delay Mechanism on EDGA." Retrieved June 30, 2020 from `https://www.sec.gov/rules/sro/cboeedga/2020/34-88261.pdf`.

U.S. Securities and Exchange Commission. 2021. "Notice of Filing of Amendments No. 3 and No. 4, and Order Granting Accelerated Approval of a Proposed Rule Change, as Modified by Amendments No. 3 and No. 4, to Introduce Periodic Auctions for the Trading of U.S. Equity Securities." March 26. SEC Approval. Retrieved July 27, 2021 from `https://www.sec.gov/rules/sro/cboebyx/2021/34-91423.pdf`.

Zhu, Haoxiang. 2014. "Do Dark Pools Harm Price Discovery?" *The Review of Financial Studies*, 27(3): 747–789.