



## Management Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Duplicated Orders, Swift Cancellations, and Fast Market Making in Fragmented Markets

Hans Degryse, Rudy De Winne, Carole Gresse, Richard Payne

To cite this article:

Hans Degryse, Rudy De Winne, Carole Gresse, Richard Payne (2025) Duplicated Orders, Swift Cancellations, and Fast Market Making in Fragmented Markets. Management Science

Published online in Articles in Advance 20 Nov 2025

. <https://doi.org/10.1287/mnsc.2023.01789>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2025, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Duplicated Orders, Swift Cancellations, and Fast Market Making in Fragmented Markets

Hans Degryse,<sup>a</sup> Rudy De Winne,<sup>b</sup> Carole Gresse,<sup>c,\*</sup> Richard Payne<sup>d</sup>

<sup>a</sup>Centre for Economic Policy Research, Katholieke Universiteit Leuven, 3000 Leuven, Belgium; <sup>b</sup>Louvain Finance (LFIN / LIDAM), UCLouvain, 1348 Louvain-la-Neuve, Belgium; <sup>c</sup>Dauphine Recherches en Management, Centre National de la Recherche Scientifique, Université Paris Dauphine-Paris Sciences et Lettres, 75775 Paris, France; <sup>d</sup>Bayes Business School, City St George's, University of London, London EC1Y 8TZ, United Kingdom

\*Corresponding author

Contact: [hans.degryse@kuleuven.be](mailto:hans.degryse@kuleuven.be),  <https://orcid.org/0000-0002-0130-057X> (HD); [rudy.dewinne@uclouvain.be](mailto:rudy.dewinne@uclouvain.be),  <https://orcid.org/0000-0001-7277-7132> (RDW); [carole.gresse@dauphine.psl.eu](mailto:carole.gresse@dauphine.psl.eu),  <https://orcid.org/0000-0003-3994-8002> (CG); [richard.payne.1@city.ac.uk](mailto:richard.payne.1@city.ac.uk),  <https://orcid.org/0000-0003-3123-5243> (RP)

Received: June 12, 2023

Revised: November 11, 2024; April 4, 2025

Accepted: April 27, 2025

Published Online in *Articles in Advance*:  
November 20, 2025

<https://doi.org/10.1287/mnsc.2023.01789>

Copyright: © 2025 INFORMS

**Abstract.** Employing unique data from 91 stocks trading on their primary exchanges and three alternative venues, we show that liquidity suppliers post duplicate limit orders on competing trading venues and cancel the duplicated orders immediately after one of them is filled. This is consistent with liquidity suppliers engaging in crossvenue market making. This duplicated-then-canceled liquidity is predominantly used by high-frequency traders when their inventories are not excessive. It reduces execution costs of fast traders on alternative venues. It, however, has some adverse impact on execution costs on primary exchanges, but those negative effects fail to outweigh the liquidity benefits of market fragmentation.

**History:** Accepted by Agostino Capponi, finance.

**Funding:** The authors thank the Agence Nationale de la Recherche [Grant ANR-15-CE33-0006] for funding the project and Centre pour la recherche économique et ses applications for granting additional research budget.

**Supplemental Material:** The data files are available at <https://doi.org/10.1287/mnsc.2023.01789>.

**Keywords:** high-frequency trading • algorithmic trading • fragmentation • limit order duplication • order cancellation

## 1. Introduction

Recent developments in equity market structure have rendered the process of identifying and implementing optimal execution and market-making strategies significantly more complex. First, the fragmentation of modern equity markets and the use of multiple trading venues by market participants mean that to understand liquidity, one must aggregate across many venues and data feeds to obtain a “consolidated” view of the market, whereas executing efficiently often requires the use of a “smart order router” (SOR) (see, for example, Foucault and Menkveld 2008). Second, the same market developments have led to changes in traders’ limit order submission strategies, which imply that “consolidated” liquidity (measured as the simple aggregate of shares available across all trading venues) is likely to overstate the actual liquidity that the average impatient market participant can access. This is because in a world of fragmented market order flow, traders implementing market-making strategies may rationally choose to place duplicate limit orders on several venues, intending for only one of those orders to execute. In this work, we empirically evaluate the extent of this order duplication, the extent to which consolidated liquidity overstates true

liquidity, and more importantly, how it is related to crossvenue market-making activity.

To illustrate the key issue, consider a simple scenario in which all participants involved in trading a stock have access to two venues. A patient investor who wishes to buy a unit of the stock might place a limit buy order on one of the two venues. She then executes if a matching market sell arrives at this venue. However, she misses out on trading opportunities if market sells are arriving at the other venue. Thus, to maximize her chances of execution, she is incentivized to place similar limit buy orders on both venues and intends, when one of the orders has executed, to cancel the other.

On the one hand, this order duplication creates a difference between true and measured liquidity. Let us imagine that an impatient but unsophisticated trader places a market sell order to hit the limit buy order posted on one of the two venues but that at the same time, the duplicate limit buy order submitted by the patient trader is executed on the other venue. If the seller’s trading technology is slower than that of the patient buyer, by the time her sell order reaches the market, the limit buy order that she targets will have been canceled. As a result, the liquidity accessible to her is less than

initially observed. We call this difference *duplicated-then-cancelled liquidity* (DCL).

On the other hand, DCL is likely to make the consolidated market more liquid. First, DCL is not necessarily inaccessible to all traders. One may fairly claim that DCL is genuine liquidity between its submission time and its cancellation time not only for fast multimarket aggressive traders but also, for the small local traders of all venues. Second, DCL may favor liquidity provision in fragmented markets by contributing to more effective inventory management for traders who operate crossvenue market-making strategies.

In brief, in a world of fragmented trading, the duplication of limit orders across venues may be a natural part of crossvenue market making, but it may also lead measured liquidity to overstate true liquidity for the average trader. The core of this paper is an attempt to quantify the importance of DCL in equity markets; to identify its determinants and in particular, how it relates to the inventories of liquidity suppliers (LSs); and to characterize its actual effects on liquidity.

It is worth noting that order duplication is not without risk. If both of the passive trader's limit buy orders in our example are hit simultaneously, she will have executed too great a quantity. This double execution may occur either because the duplicated orders are hit at each venue by two different traders or because a single trader using a smart order router intentionally and simultaneously executes the passive trader's orders on both venues. This simple example implies that the incentive to duplicate limit orders across venues is greater for traders who have a trading speed advantage over the average trader, but the incentive is weakened by the presence of smart order routers. Our analysis of the determinants of DCL takes those factors into consideration by clearly identifying trading speed levels and by building a measure of SOR activity.

To conduct our empirical work, we take advantage of a unique data set that covers 91 European stocks trading on their respective primary exchanges and the three largest alternative European trading venues for the month of May 2013. The data contain the usual order level and individual trade information that is common to many modern microstructure databases, but importantly, the data also provide anonymized information on the market members who submitted each order. Thus, we can track market members across time, across stocks, and across trading venues. This identity information can also be used to characterize those participants in terms of trading speed and technology.

With these data, we measure DCL by computing a trader's voluntary cancellations of liquidity on one venue following execution of one of that trader's similar orders on another venue. Then, we aggregate across traders, venues, and time to assess the overall size of DCL. We then regress DCL measures on a set of trader

characteristics, venue characteristics, and exogenous variables to characterize the determinants of duplication. Last, we analyze the impact of DCL on the trading costs of slow and fast liquidity takers (LTs).

We find that DCL is an important driver of order cancellation activity. To a first approximation, execution of one of the average participant's limit orders on a particular venue leads her to cancel a quantity equivalent to roughly 19% of the size of that trade on each venue where she has posted similar orders. When consolidating across all venues where the participant has posted limit orders, the average DCL rises to 59% of the initial traded quantity. High-frequency traders (HFTs) and traders acting as a principal rather than as an agent have the largest measures of DCL. This supports the idea that DCL is generated, at least in part, by high-speed market-making strategies. We proceed to a Tobit analysis of the determinants of DCL based on data measured at a 15-minute interval. We find that in addition to the results above, liquidity suppliers are marginally less likely to duplicate orders when their inventory level is large. This leads us to reject the hypothesis that duplication is used as a tool for rebalancing extreme inventories. Instead, it appears that traders are more comfortable submitting orders to multiple venues when inventories are not excessive. More explicitly, this indicates that limit order traders implementing liquidity-supplying strategies are more likely to use DCL in the first phase of those strategies (i.e., when they build trading positions) rather than in the second phase (i.e., when they unwind them). We also find that when the prevalence of smart order routing is particularly large, it tends to reduce DCL. This result is in line with the intuition presented earlier and the likelihood of multiple executions when smart order routers are a significant factor in the market.

Finally, we present an analysis of the implications of DCL, focusing on the DCL activity generated by HFTs. We examine whether the level of duplication impacts the execution costs measured by effective spreads that various trader groups pay. We find that DCL on primary venues leads to increases in slow trader and algorithmic trader (AT) execution costs. The effect, however, is economically and statistically small for slow traders. On the other hand, there is strong evidence that the DCL activity of HFTs reduces the trading costs of algorithmic traders on alternative trading venues. Finally, DCL reduces HFT trading costs on all venues. We interpret this as implying that DCL forms an important part of HFT market-making strategies, which allows their counterparties access to greater liquidity on nonprimary trading venues.

Overall, our results indicate that order duplication is used by fast trading firms employing crossvenue market-making strategies in fragmented markets. This interpretation is supported by the results that

duplication is more heavily used by those trading as principal and by those who predominantly supply liquidity and that duplication is employed when trader inventory levels are not excessive. Further support comes from the result that duplication is greater in more fragmented markets and that it is smaller when smart order routing is prevalent. Overall, order duplication contributes to the formation of liquidity on all venues, with greater benefits on alternative venues. Its incidence does, however, raise questions about the use of simple consolidated liquidity measures to assess market quality. Indeed, we find that higher DCL is associated with greater execution costs for less sophisticated traders. Yet, the low economic significance of this effect and the limited crossmarket scale of DCL—around 4% of total crossmarket consolidated depth—mean that this positive effect on execution costs is unlikely to be sufficiently great to challenge the benefits of fragmentation found in earlier work.

The rest of the paper is structured as follows. Section 2 contains a brief overview of relevant literature. Section 3 is an introduction to our data. Section 4 presents our measurements of DCL. Section 5 gives a description of how we classify market participants using our data and provides DCL statistics by member category. Section 6 contains our analysis of the determinants of DCL. We examine the impact of DCL on trading costs in Section 7, and Section 8 provides some conclusions from our work.

## 2. Literature Review and Research Objectives

We are interested in measuring and characterizing the determinants of duplicated-then-canceled liquidity. By DCL, we mean liquidity that is supplied to markets but that is not intended to execute in full. The strand of the literature that is closest to our focus is that which investigates order cancellations related to liquidity oversupply, both in a single-market setting and in a multimarket setting.

### 2.1. Literature on Liquidity Oversupply and Order Cancellations

DCL may occur in a single consolidated market, with a trader submitting multiple buy or sell orders to different levels of an order book (in order to gain time priority), only one of which is intended to execute. Recent work has demonstrated that there may be oversupply of depth on a single venue resulting from the imposition of time priority and variations of trading speed across participants. Yueshen (2025), for example, argues that following changes in asset prices, there may be a race by fast traders to be the first in line at the new equilibrium price, leading to a temporary spike in depth before traders realize their actual position in the queue and through subsequent cancellations, depth normalizes. Baruch and Glosten (2013) show that

flickering or fleeting quotes (i.e., quoting of limit orders and their quick cancellation) can be an equilibrium strategy in a trading game on one venue. Limit order traders then manage their undercutting exposure by rapidly canceling their quotes and replacing them with new randomly chosen ones. This behavior is often associated with HFTs given their speed advantage.

Blocher et al. (2016) identify clusters of extremely high and extremely low limit order cancellation activity using data on all of the S&P 500 stocks for the calendar year of 2012. They find that cancel clusters largely appear to be generated by HFTs sparring with one another to get to the front of the limit order queue rather than HFTs trapping unsuspecting investors into bad executions. Dahlström et al. (2024) investigate the economic rationale behind limit order cancellations from the perspective of liquidity suppliers. They show that changes in common values affect the value of a limit order depending upon the queue position, but HFTs behave in a similar way to other traders. These papers suggest that competition between fast traders on the same venue can lead to “excess” depth in the short run that is eliminated by cancellation activity. Dahlström et al. (2024) further show that trades at competing venues lead to significant cancellations at the primary venue; the economic significance of this force relative to other determinants of cancellations, however, is low.

DCL may also arise because of fragmentation in trading across venues, and there lies our research interest. Traders who are connected to many competing trading venues can benefit by accessing the separate liquidity pools on those venues. Chen and Duffie (2021) show that market fragmentation reduces market depth on each exchange but leads to more aggressive overall order submission. Empirical research indicates a strong link between fragmentation and measured liquidity. Foucault and Menkveld (2008) show that because of the absence of time priority across markets, consolidated depth is larger after the entry of a new order book. O’Hara and Ye (2011) find that for U.S. stocks, spreads are tighter and price efficiency is higher with fragmentation. Degryse et al. (2015) find that lit fragmentation (i.e., fragmentation across pretrade transparent venues) in Dutch stocks has increased liquidity through reductions in bid-ask spreads and increases in depth across markets. Gresse (2017) employs data for stocks listed on the London Stock Exchange (LSE) and Euronext and finds that lit fragmentation improves bid-ask spreads and depth across markets. We will add to this literature by showing that in fragmented markets, individual traders place duplicate orders on several exchanges. This challenges the result that fragmentation leads to larger *measured*, *consolidated* liquidity as with duplicated limit orders, *measured* and *real* liquidity may differ.

In a multimarket setting where some liquidity suppliers have a latency advantage, a reasonable market-

making strategy may involve the posting of duplicate limit orders on more than one venue, intending for only one of the orders to execute and (partially) canceling the duplicates once an execution occurs. The latency advantage enjoyed by these traders who run market-making strategies means that they face limited asymmetric information risk and that the risk of being overfilled is small. It is this order duplication across venues that we define as DCL and that implies that measured, consolidated liquidity is larger than real liquidity. It is worth noting, however, that to the extent that it makes inventory management easier for those operating market-making strategies, it is likely to improve real liquidity.

Our work is also related to other papers modeling the behavior of investors when markets are fragmented. Baldauf and Mollner (2021), for example, model trading in fragmented markets where HFTs are liquidity providers or arbitrageurs and where other investors are liquidity demanders. Fragmentation has two effects. First, fragmentation intensifies competition on trading fees. Second, it creates an exposure channel as HFT standing orders can now be sniped on several venues by other fast traders. Malinova and Park (2020) also study “sniping” behavior by investigating the reactions of market participants following trades that simultaneously hit standing orders on multiple markets. Employing proprietary, trader-level data covering multiple equity trading venues in Canada, they document aggressive market reactions following those multimarket trades; fast traders initiate “follower-trades” on the same side of the market, and market makers race to avoid trading against these “sniping” orders by canceling standing limit orders. Our paper contributes to this line of work by studying how the liquidity provision of traders with different speeds on a venue is modified when their standing orders are executed on other venues.

From this literature review, our research agenda has three steps: (1) build an empirical measure of DCL that best approximates the actual magnitude of DCL in today’s markets, (2) identify the factors determining DCL as well as the role of DCL in crossvenue market making, and (3) measure the impact of DCL on the trading costs of liquidity takers. We now develop a set of hypotheses to guide the empirical work related to research goal (2).

## 2.2. Testable Hypotheses

The fundamental idea underlying our analysis is that when the order flow in a stock is fragmented across several order books, optimal market-making strategies will likely entail posting duplicated orders on multiple venues. In this scenario, cancellations associated with duplicated orders are frictions associated with this market-making activity.

To develop this idea, a basic high-speed market-making strategy in an order book requires first, the

placing of limit orders on one side of the book (phase 1) and second, after execution of one of those orders, unwinding the position by placing limit orders on the other side (phase 2). Thus, one earns the bid-ask spread. It is valuable if the whole strategy is completed within a very short time interval so as to minimize risk and to enable repetition of the strategy as many times as possible within a day. Execution speed is thus key to maximizing the profits expected from such strategies. In fragmented markets, a trader operating a high-speed market-making strategy cannot know with certainty on which venue liquidity traders will appear first. Therefore, duplicating limit orders across books with the intention to cancel residual orders as soon as the desired quantity is executed in one book (i.e., DCL) increases expected market-making profits by reducing both execution delays and nonexecution risk.

**Hypothesis 1.** *This improvement in execution speed and probability is effective if marketable orders actually arrive on several venues (i.e., if the order flow is sufficiently fragmented). Thus, we expect DCL to increase with fragmentation.*

**Hypothesis 2.** *The incentive to post duplicated orders and the resulting DCL are greater when other options to improve execution probability, such as competing on price, are not available. DCL should then be greater when the tick size is more likely to be a binding constraint on price competition (i.e., when a large tick size makes price undercutting expensive or impossible) and market-making incentives to capture rents are greater (an argument similar to that in Yueshen 2021). For that reason, we expect DCL to increase in the relative tick size.*

**Hypothesis 3.** *As duplicated liquidity is a tool used to increase the profits of limit order traders in their market-making activities, we expect frequent liquidity suppliers to post more DCL than otherwise similar traders.*

**Hypothesis 4.** *As duplicated liquidity is a tool used to increase the profits of limit order traders in their market-making activities, we expect traders acting as principal to post more DCL than otherwise similar traders.*

**Hypothesis 5.** *The eagerness of a liquidity supplier to trade may also depend on her pretrade inventory level. It is possible that a liquidity supplier might use DCL in an attempt to quickly reduce inventory when those inventory levels deviate significantly from the optimal level (i.e., in the inventory management phase of market making). An alternative view is that DCL is more extensively used when inventory is low and when posting duplicated liquidity carries little risk (i.e., in the inventory-building part of a market-making strategy). Ex ante, it is unclear which, if either, of these mechanisms might be employed. Thus, we can only hypothesize that the DCL posted by a market member is related to the deviation between her stock inventory and its*

normal level, and the sign of the estimated inventory coefficient will tell us which of the preceding conjectures is consistent with the data.

**Hypothesis 6.** *Fast traders are more likely to be actively making markets than slow traders through duplicated orders. To handle the possible friction of overtrading inherently linked to market-making strategies (i.e., the risk of being executed at multiple locations such that total quantity traded exceeds desired quantity), we expect the DCL of a market member to increase with their trading speed advantage.*

That is, we expect fast traders to exhibit greater DCL than slow traders. Their trading speed advantage also depends on the technology used by those who they are trading against. In particular, the trading speed advantage that they use for fast cancellations will not be effective if, on the other side of the market, sophisticated market order traders use SORs to hit limit orders on several platforms simultaneously as modeled in Chen and Duffie (2021).<sup>1</sup>

**Hypothesis 7.** *We thus posit that DCL decreases with the presence of SORs.*

**Hypothesis 8.** *Finally, trading speed advantages are better exploited on platforms with lower latency, a feature shared by the alternative venues in our study as reported by Menkveld (2013) and Gresse (2014). This leads us to expect DCL to be greater on alternative platforms.*

Testing those eight hypotheses allows us to identify several economic drivers of DCL and thus, to shed light on how crossvenue market-making strategies operate. Following this, the other major contribution of our work is to assess the impact of DCL on the execution costs of different groups of traders.

### 3. Sample, Data, and Market Organization

We employ a proprietary data set collected by the European Securities and Markets Association (ESMA) and several national competent authorities for the month of May 2013. It consists of 91 stocks that are primary listed on the historically main exchanges of nine countries, these being Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, and the United Kingdom. Those national exchanges will be referred to as “primary” exchanges and denoted *PE* in the empirical analysis.<sup>2</sup> The data set covers trading and quoting activity on the primary exchanges and on the three largest alternative exchanges on which the sample stocks were admitted for trading at that time, namely BATS, Chi-X, and Turquoise.<sup>3</sup> These three venues will be referred to as “alternative” exchanges and denoted *ALT* in the rest of the paper. Together with the primary exchanges, they deal with the vast majority of trading activity for each stock.<sup>4</sup> Bouveret et al. (2014) were the first to employ this data set in their analysis of the extent of

HFT in European stock markets. Further details on the construction and content of the data set can be found there.

All exchanges in our study are regulated under the Markets in Financial Instruments Directive (MiFID). Key provisions in MiFID include the abolition of trading concentration on main exchanges, pretrade and posttrade transparency requirements for all trading venues, and best-execution rules. A key difference between MiFID and Regulation National Market System (Reg NMS) in the United States is that MiFID defines best execution as not only a matter of the price of a trade but also, “costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order” (Article 21). Although Rule 602b of Reg NMS obliges trading venues to execute orders at the best price or to reroute them to the venue quoting the best price to prevent trade throughs, MiFID has softer requirements. Consequently, European exchanges do not route orders to each other, and order flow fragmentation might distort both time priority and occasionally, price priority.

In terms of market organization, all trading platforms considered in our study operate as open, transparent, and anonymous electronic order books on which buy and sell orders are continuously matched from the open to the close according to price/time priority rules. Primary exchanges commence and finish their trading sessions with call auctions, whereas no call auctions are organized on alternative venues either at the open or at the close. Further, alternative venues use a make/take fee structure that remunerates liquidity-providing orders and charges aggressive orders.

The set of stocks in the data set was built using a stratified sampling approach taking into consideration market capitalization, value traded, and fragmentation. For each country, stocks were split by quartiles according to their market value, value traded, and level of fragmentation across venues using December 2012 data. A random draw was performed to select stocks in each quartile. To account for the relative size of the markets, greater weight was put on larger countries, but at least five different stocks were selected from each country. This procedure yielded an original sample of 100 stocks, from which 9 stocks had to be excluded because of thin trading issues.<sup>5</sup> As a result, we have only four stocks in two of our sample countries. The final sample includes stocks with very different features. The average daily value traded ranged from less than EUR 0.1 million (mn) to EUR 611 mn. In terms of market capitalization, values ranged from EUR 18 mn to EUR 122 billion. The breakdown of stocks per country and descriptive statistics for those stocks are provided in Table 1.

The entire data set includes around 10.5 million trades and 456 million messages. Message types include transactions plus order entries, modifications, and cancellations. The unique feature of the data set is

**Table 1.** Descriptive Statistics on Sampled Stocks

Country	No. of stocks	Mean, min, and max	Market value (EUR mn)	Value traded (EUR mn)	Crossmarket bid-ask spread, %	Market share of the primary exchange, %
Belgium	6	Mean	24,327	2,012	0.0465	72.13
		Min	843	86	0.0181	62.44
		Max	118,942	8,134	0.0956	88.78
France	15	Mean	7,957	1,632	0.0362	74.73
		Min	195	2	0.0063	62.35
		Max	55,979	12,658	0.1006	97.30
Germany	13	Mean	10,039	1,997	0.0962	74.80
		Min	242	10	0.0084	59.25
		Max	71,713	15,074	0.4480	95.63
Ireland	4	Mean	4,551	291	0.0450	86.20
		Min	1,599	46	0.0010	79.97
		Max	7,898	709	0.0951	93.07
Italy	11	Mean	6,495	1,454	0.0305	86.84
		Min	292	7	0.0015	79.01
		Max	27,628	6,234	0.1609	98.31
Portugal	4	Mean	6,035	944	0.0047	74.92
		Min	2,080	612	0.0010	63.14
		Max	10,857	1,090	0.0135	85.44
Spain	12	Mean	9,650	1,884	0.0098	85.02
		Min	801	299	0.0024	78.77
		Max	40,712	10,613	0.0238	92.35
The Netherlands	11	Mean	7,747	1,771	0.0181	75.43
		Min	383	54	0.0014	64.80
		Max	50,233	9,036	0.0607	87.64
The United Kingdom	15	Mean	8,529	1,228	0.0189	65.27
		Min	395	16	0.0028	53.47
		Max	69,843	6,969	0.0480	79.80
Total	91	Mean	9,481	1,468	0.0340	77.26
		Min	195	2	0.0010	53.47
		Max	118,942	15,074	0.4480	98.31

*Notes.* This table reports the number of stocks sampled by country and for each country, the average, minimum, and maximum values of the market value in millions of euros; the total traded value in May 2013 in millions of euros; the crossmarket bid-ask spread; and the market share of the primary exchange. Four markets are considered: the primary exchange, Chi-X, Bats, and Turquoise.

that it contains information on the identity of the market participant behind each message, allowing us (i) to follow a market participant across trading venues and (ii) to categorize each participant as an HFT or a non-HFT. There is also a capacity flag for each event that indicates whether the member in question is acting in a proprietary or agency capacity. We provide further details on the classification of members in Section 5.

#### 4. Assessing the Level of Duplicated-then-Canceled Liquidity

Our definition of DCL above suggests an approach of measuring order duplication by identifying order cancellations on one venue in response to trades on another.<sup>6</sup> However, van Kervel (2015) argues that one might also observe such crossvenue cancellations in

response to trades on other venues, even in settings with no DCL (by our definition). van Kervel (2015) builds a model with multiple venues in which HFT market makers post quotes on all venues simultaneously. In the absence of any new information, those HFTs would be willing to trade at those quotes on all venues and would not choose, for example, to cancel or modify quotes on venue B in response to a trade on venue A. In this sense, those quotes reflect real depth and not duplicated orders. However, if there is asymmetric information, then a trade on venue A will lead to quote updating through cancellations and modifications on all other venues. Again, this is not a result of order duplication, but it is the rational updating of quotes in response to new information contained in the trade on venue A. Thus, one observes crossvenue cancels in a world without DCL. Employing data from the

LSE and four competing exchanges, van Kervel (2015) finds that once a market order consumes liquidity on one venue, the depth available at other venues is reduced. Two takeaways from the work of van Kervel (2015) are that (1) it is important for us to account for asymmetric information effects if we want to understand cancellation activity and that (2) estimates of DCL simply based on cancellations without tracking traders individually would be biased as those cancellations might reflect the rational updating of dealers' quotes in response to information revealed by trades. Our research overcomes this identification challenge by following the same traders across venues. It is worth noting that previous research could not track individual traders across venues in their data.

Section 4.1 describes the methodology that we use to measure DCL and to aggregate it at different levels. Section 4.2 describes how we check whether the DCL that we measure is a genuine reduction in depth or whether it is immediately followed by resupply of liquidity by the same trader on the same side of the market but at a different price point, thereby suggesting information-based quote updating. In Section 4.3, we describe how we investigate whether DCL is immediately followed by liquidity provision by the same trader on the opposite side of the market, reflecting market-making activity. Section 4.4 then reports descriptive statistics regarding the previous steps.

#### 4.1. Measuring DCL

Our DCL metric is based on the following simple intuition. Assume that a trader is posting limit sell orders, for example, on several venues simultaneously. Also, assume that at a certain time, the limit order that they posted on the first venue is executed. If after the execution of the order on the first venue, the trader's limit orders on other venues are left in their respective order books, then those orders constitute real liquidity. If, on the other hand, when the order on the first venue executes, the limit orders on other venues are swiftly canceled, then those canceled orders represented DCL.

As the simple example above makes clear, DCL has several dimensions. It is trader specific, and it might be venue specific. Also, there are several parameters to be specified. How quickly does a trader's order have to be canceled in response to the execution of another of that trader's orders on a different venue to qualify as DCL? How similar does the canceled order have to be to the executed order to count as DCL? Any definition of DCL needs to be flexible enough to take account of all of the above.

We begin with a specification of DCL as follows. Assume that at time  $\tau$ , a limit sell order posted by member  $m$  for stock  $i$  was executed on venue  $tv$ , the trade venue, and that member  $m$  had also posted a limit sell

order for stock  $i$  on venue  $qv$ , the quote venue. Then, the sell-side DCL posted by  $m$  on venue  $qv$  is measured by the following quantity:

$$\begin{aligned} & DCLQTY_{tv \rightarrow qv}^{ask}(\tau; \Delta\tau; i; m) \\ &= PREQTY_{qv}^{ask}(\tau; i; m) - POSTQTY_{qv}^{ask}(\tau; \Delta\tau; i; m) \\ &\quad - \sum_{\tau; \Delta\tau} Volume_{qv}^{buy}(i; m), \end{aligned} \quad (1)$$

where  $PREQTY_{qv}^{ask}(\tau; i; m)$  is the total limit sell order quantity posted by trader  $m$  on venue  $qv$  at the last order book snapshot prior to the trade executed on venue  $tv$  and  $POSTQTY_{qv}^{ask}(\tau; \Delta\tau; i; m)$  is the total limit sell order quantity attributable to member  $m$  on venue  $qv$  at the order book snapshot that is exactly  $\Delta\tau$  seconds after the original snapshot. Thus, the first pair of terms on the right-hand side of the definition measures the reduction in quantity posted by trader  $m$  on venue  $qv$  over a small time window (i.e.,  $\Delta\tau$ ) around the time of the trade on venue  $tv$ . The final term on the right-hand side consists of all executions against trader  $m$ 's limit sell orders on venue  $qv$  in that same window.  $Volume_{qv}^{buy}(i; m)$  is defined as the size of a market buy order, executing against one of market member  $m$ 's orders on venue  $qv$  for stock  $i$  at any time within the time window. Therefore, all that this definition does is to take the change in total quantity offered by trader  $m$  and deduct that part of the change that is because of execution activity. The remainder represents voluntary reduction in limit order provision on venue  $qv$  after the trade on venue  $tv$ , and we count this as DCL.

As order book snapshots were built every 10 milliseconds (ms) in the database, the time interval over which we build this measure is always a multiple of 10 ms. In our baseline specifications, we set the interval to be exactly 10 ms but do some robustness analysis using longer windows.<sup>7</sup> The facts that our order book data are on a 10-ms sampling frequency and that trades are sampled more frequently also mean that there will be some noise in our DCL measure. Assume that we are measuring DCL precisely over a 10-ms interval. A trade arriving just after an order book snapshot will see the majority of this 10-ms interval coming after the trade, whereas a trade arriving just before an order book update will have most of the 10-ms interval pretrade. Thus, although in this example, depth changes are always measured over a 10-ms interval, there will be small variations across trades in the portion of that interval that comes before the trade and the portion that comes afterward.

In the definition above, depth measures  $PREQTY_{qv}^{ask}(\cdot)$  and  $POSTQTY_{qv}^{ask}(\cdot)$  are quantities available in the order book of venue  $qv$  that lie in a range specific to each stock around the midquote. We set that range as follows. First, we construct the time series of differences between the third most competitively priced limit buy and limit sell

orders from the consolidated order book. Then, we compute the 90th percentile of the distribution of those differences. At each point where we compute DCL for a stock, that 90th percentile value is used to define a symmetric band centered on the current midquote, and only orders that are inside this band are counted toward DCL. We have chosen the width of the band to ensure that our DCL measure captures a majority of order activity, focusing on cancellations of orders with prices close to the execution price on  $tv$  while excluding orders that lie a long way from the stock's midquote.

The baseline DCL measure above, which is expressed in number of shares, is trader, trade time, stock, venue, and side specific, and we want to aggregate these data so that they can be compared across stocks and times. To make the data comparable across stocks and to aggregate up to the daily level, we use two relative measures of DCL, which we define precisely below: (1) a DCL measure that scales by trade size and (2) a DCL measure that scales by depth. These two different scalings of DCL allow us to focus on different research questions. We consider DCL scaled by trade size to be appropriate for assessing how DCL is used by members in their individual trading strategies, particularly for members who actively supply liquidity in multiple markets. For that reason, it will be our main variable of interest in our investigation of DCL determinants. DCL scaled by depth is of greater interest for assessing the overall crossmarket magnitude of DCL, and it will be our main variable of interest when measuring the impact of DCL on trading costs at the market-wide level.

**4.1.1. DCL Relative to Trade Size.** Our first approach to scale DCL is to divide it by the size of the original trade on venue  $tv$ . This allows us to ask, for example, if a trade on one venue leads to the removal of a similarly sized order on another venue. We, therefore, construct in the following equation a DCL measure where the numerator aggregates DCL quantities for a given day, a given stock, a given member, and a given quote venue and the denominator of the computation aggregates the sizes of the trades of a given trade venue, which triggered the DCL measurements appearing at the numerator:

$$DCL_{tv \rightarrow qv}^{tradesize}(\Delta\tau; i; d; m) = \frac{\sum_{s=1}^{S_d^{tv}} DCLQTY_{tv \rightarrow qv}^{bid}(\tau_s; \Delta\tau; i; m) + \sum_{b=1}^{B_d^{tv}} DCLQTY_{tv \rightarrow qv}^{ask}(\tau_b; \Delta\tau; i; m)}{\sum_{s=1}^{S_d^{tv}} TradeSize_{tv}^{sell}(\tau_s; i; m) + \sum_{b=1}^{B_d^{tv}} TradeSize_{tv}^{buy}(\tau_b; i; m)} \quad (2)$$

In Equation (2),  $DCLQTY_{tv \rightarrow qv}^{bid}(\tau_s; \Delta\tau; i; m)$  and  $DCLQTY_{tv \rightarrow qv}^{ask}(\tau_b; \Delta\tau; i; m)$  are DCL quantities computed as in Equation (1).  $TradeSize_{tv}^{sell}(\tau_s; i; m)$  ( $TradeSize_{tv}^{buy}(\tau_b; i; m)$ ) is the size of

the  $s$ th sell trade ( $b$ th buy trade) executed against a bid (ask) limit order of member  $m$  on venue  $tv$  at time  $\tau_s$  ( $\tau_b$ ), which triggered the measurement of  $DCLQTY_{tv \rightarrow qv}^{bid}(\tau_s; \Delta\tau; i; m)$  ( $DCLQTY_{tv \rightarrow qv}^{ask}(\tau_b; \Delta\tau; i; m)$ ) on venue  $qv$  on day  $d$  for stock  $i$ .  $\tau_s$  and  $\tau_b$  denote the respective times of the  $s$ th sell trade and the  $b$ th buy trade on venue  $tv$ , which triggered a DCL measurement for member  $m$  on venue  $qv$  on day  $d$  for stock  $i$ , whereas  $S_d^{tv}$  and  $B_d^{tv}$  are the respective total numbers of those trades. All other items of notation are similar to those used in Equation (1).

**4.1.2. DCL Relative to Depth.** Alternatively, we scale DCL by pretrade quantities displayed in the book in order to estimate at market level the fraction of depth that is duplicated and likely to be swiftly canceled following executions of duplicates on another venue. To that end, we construct an alternative DCL measure where in the denominator of the computation, we replace the size of the trade that triggered the DCL measurement with the pretrade depth contributed by member  $m$  on venue  $qv$ , such as in the following equation:

$$DCL_{tv \rightarrow qv}^{depth}(\Delta\tau; i; d; m) = \frac{\sum_{s=1}^{S_d^{tv}} DCLQTY_{tv \rightarrow qv}^{bid}(\tau_s; \Delta\tau; i; m) + \sum_{b=1}^{B_d^{tv}} DCLQTY_{tv \rightarrow qv}^{ask}(\tau_b; \Delta\tau; i; m)}{\sum_{s=1}^{S_d^{tv}} PREQTY_{qv}^{bid}(\tau_s; i; m) + \sum_{b=1}^{B_d^{tv}} PREQTY_{qv}^{ask}(\tau_b; i; m)} \quad (3)$$

On the right-hand side of Equation (3),  $DCLQTY_{tv \rightarrow qv}(\cdot)$  and  $PREQTY_{qv}(\cdot)$  are summed for all trades within a given day to give aggregated DCL for member  $m$  on venue  $qv$  in response to executions on venue  $tv$  on day  $d$  for stock  $i$ . All items of notation are similar to those used in Equations (1) and (2).

## 4.2. Measuring Order Book Refilling in the 10 ms Immediately After Duplicated Order Cancellations

One may argue that our DCL measure is not necessarily capturing the cancellation of duplicated limit orders that traders have posted to optimize execution probabilities but that it could reflect quote updates in reaction to information contained in trades on other venues. If these quote updates are because of orders being repriced, we should observe order cancellations<sup>8</sup> and then, swift resubmissions at different prices but for roughly the same quantity in the quote venue's order book. No such resubmissions should occur in the case of genuine DCL. Thus, to distinguish DCL from quote updating, we compute a book refill rate for the 10 ms after the time window over which DCL is measured. For a given member whose order cancellation has contributed to

our DCL calculation, this refill rate equals the liquidity added by that same member on the same venue where DCL is being measured.<sup>9</sup> To be explicit about the calculation of the refill rate, let us return to the example that we used when discussing the DCL calculation in Equation (1). At time  $t$ , a limit sell order submitted by member  $m$  is executed on venue  $tv$  for stock  $i$ . At the same time,  $m$  also has limit sell orders posted on venue  $qv$  for stock  $i$ . We measure the sell-side DCL of  $m$  on venue  $qv$  by looking at her cancellations inside a 10-ms time window that starts at the closest 10-ms time stamp preceding trade time  $\tau$ . The refill rate is calculated over the next 10-ms window in the following way:

$$\begin{aligned} & \text{Refill}_{tv \rightarrow qv}^{\text{ask}}(\tau + 10 \text{ ms}; i; m) \\ &= \left( \begin{array}{c} \text{POSTQTY}_{qv}^{\text{ask}}(\tau + 10 \text{ ms}; i; m) \\ -\text{PREQTY}_{qv}^{\text{ask}}(\tau + 10 \text{ ms}; i; m) \\ + \sum_{\tau+10 \text{ ms}} \text{Volume}_{qv}^{\text{buy}}(i; m) \end{array} \right) / \\ & \text{DCLQTY}_{tv \rightarrow qv}^{\text{ask}}(\tau; 10 \text{ ms}; i; m), \end{aligned} \quad (4)$$

where  $\text{PREQTY}_{qv}^{\text{ask}}(\tau + 10 \text{ ms}; i; m)$  is the total limit sell order quantity posted by trader  $m$  on venue  $qv$  at the first 10-ms order book snapshot following trade time  $\tau$  (on venue  $tv$ ) and  $\text{POSTQTY}_{qv}^{\text{ask}}(\tau + 10 \text{ ms}; i; m)$  is the total limit sell order quantity posted by member  $m$  on that same venue 10 ms later.  $\sum_{\tau+10 \text{ ms}} \text{Volume}_{qv}^{\text{buy}}(i; m)$  consists of all executions against trader  $m$ 's limit sell orders on  $qv$  in that same 10-ms window starting after the initial trade. When the volume measure is added to the difference in quantities, it yields the amount of liquidity that member  $m$  has added to the quote venue book immediately after the duplicated order cancellations. This is then expressed as a percentage of the 10-ms DCL measured for the same trade and the same member  $m$  on the quote venue. A positive refill rate indicates that members refill the book after canceling orders, whereas a negative refill rate indicates that the members continued canceling liquidity after the end of the DCL window. Those refill rates are computed for all trades that generated positive DCL and are then averaged across time, members, and stocks by countries, platforms, stock terciles, and member categories.

### 4.3. Measuring Opposite-Side Liquidity Provision in the 10 ms Immediately After Duplicated Order Cancellations

In the same spirit, we look at whether DCL is followed by liquidity provision by the same member  $m$  over the next 10 ms on the same venue  $qv$  but on the opposite side of the market. We do so to check whether DCL could signal the turning point of a market-making strategy. In other words, the cancellations that underlie the DCL measurement might occur when the trader stops providing liquidity on one side and switches to the

other side of the market. Returning to the example used to illustrate Equations (1) and (3), after measuring sell-side DCL by member  $m$  on venue  $qv$  over a given 10-ms time window, we build a measure of liquidity provision by the same member  $m$  on the buy side of venue  $qv$  over the next 10-ms window as follows:

$$\begin{aligned} & \text{LiqProv}_{tv \rightarrow qv}^{\text{bid}}(\tau + 10 \text{ ms}; i; m) \\ &= \left( \begin{array}{c} \text{POSTQTY}_{qv}^{\text{bid}}(\tau + 10 \text{ ms}; i; m) \\ -\text{PREQTY}_{qv}^{\text{bid}}(\tau + 10 \text{ ms}; i; m) \\ + \sum_{\tau+10 \text{ ms}} \text{Volume}_{qv}^{\text{sell}}(i; m) \end{array} \right) / \\ & \text{DCLQTY}_{tv \rightarrow qv}^{\text{ask}}(\tau; 10 \text{ ms}; i; m), \end{aligned} \quad (5)$$

where  $\text{LiqProv}_{tv \rightarrow qv}^{\text{bid}}(\tau + 10 \text{ ms}; i; m)$  is the depth added by member  $m$  on the bid side of  $qv$  in the 10 ms following the cancellation of duplicated limit orders on the ask side for a quantity of  $\text{DCLQTY}_{tv \rightarrow qv}^{\text{ask}}(\tau; 10 \text{ ms}; i; m)$ . The measure is then expressed as a percentage of  $\text{DCLQTY}_{tv \rightarrow qv}^{\text{ask}}(\tau; 10 \text{ ms}; i; m)$ . We refer to this measure as the opposite-side liquidity provision rate. Opposite-side liquidity provision rates are computed for all trades that generated positive DCL and are then averaged across time, members and stocks by countries, platforms, and pairs of platforms.

### 4.4. Descriptive Statistics for DCL

We present several descriptive statistics to assess the magnitude of DCL at the consolidated market level and to understand whether this magnitude differs across stocks and trading venues. Those statistics are established for both DCL scaled by trade size and DCL scaled by depth by averaging the measures of Equations (2) and (3) across stocks, members, pairs of venues, and venues according to the following steps.

Step 1. We aggregate DCL across members for a given stock on a given day for a given pair of venues. Although this step is relatively straightforward for DCL as a fraction of trade size, for an aggregation of DCL as a fraction of depth across members on a given day for a given stock, we construct a weighted average DCL, where the weight for member  $m$  is equal to the average contribution of that member to the depth of stock  $i$  on the quote venue considered ( $qv$ ) over the day considered.

Step 2. We compute monthly averages of daily mean DCL for each stock and each pair of venues.

Step 3. We compute cross-stock averages of DCL per pair of venues. For each pair of venues, this average reflects the mean level of DCL on the quote venue ( $qv$ ) observed because of executions on the trade venue ( $tv$ ).

Step 4. We compute a single number to summarize the scale of the DCL issue on a single venue. This entails averaging across trade venues to focus on a single quote venue. The weight used in this averaging for

venue  $tv$  is equal to the total volume executed on  $tv$  over the sample divided by the sum of the volumes on all three trade venues.

Step 5. We conduct a last aggregation of DCL across venues at the consolidated market level.

Table 2 presents the average of DCL as a percentage of trade size for DCL measurements within four different time windows ranging from 10 to 100 ms.<sup>10</sup> Those figures measure how many shares a passive trader cancels on one venue, on average, after being executed for 100 shares on another venue. Panel A of Table 2 reports DCL means across all stocks, all member categories, and all trading venues. Panel B of Table 2 reports DCL mean levels by platform by taking for a given quote venue (where DCL is measured) the average DCL weighted across the trades that trigger our measurements. Panel C of Table 2 breaks down DCL by pairs of venues. The first column in Table 2 gives the name of the venue where DCL is being measured (the quote venue) and the name of the venue where the trade that triggers the measurement occurred (the trade venue). For example, a trade of 100 shares on Chi-X leads the passive trader of that trade to cancel outstanding limit

orders on the primary exchange for a quantity of nearly 23 shares on average within a 10-ms window.

According to Table 2, at the 10-ms horizon, the average level of DCL as a proportion of trade size is 19% on average across all stocks but reaches more than 30% in some cases. Thus, for example, after a trade on Turquoise, one subsequently sees nearly 31% of the trade quantity canceled on BATS by the same trader. The difference across venues in average DCL as a fraction of trade size is fairly small with, if anything, DCL being larger on the primary exchange. Looking at pair-wise average DCL levels, it is clear that DCL is at its lowest level on alternative venues (between 13% and 17%) when the triggering trade occurs on the primary exchange, whereas DCL reaches its highest (around 30%) on BATS and Chi-X when the triggering trade is on a different alternative venue.

DCL means established at longer horizons, although slightly increasing with the length of the time window, remain similar in magnitude (21% at the 100-ms horizon across all stocks). With regard to the differences between platforms and pairs of platforms, the same patterns are observed for all time windows.

**Table 2.** Average Level of DCL as a Percentage of Trade Size

	10 ms, %	20 ms, %	50 ms, %	100 ms, %
Panel A: All platforms				
All platforms	18.89	20.11	20.34	21.07
Panel B: By platform				
Primary exchange	20.54	21.83	21.61	22.31
Chi-X	18.26	19.54	21.35	22.14
Turquoise	18.62	19.63	20.34	21.54
BATS	16.82	18.41	20.74	21.31
Panel C: By pair of platforms (quote venue and trade venue)				
Primary exchange				
Chi-X	22.87	24.11	22.61	23.49
BATS	19.02	21.32	21.91	22.28
Turquoise	19.43	20.16	21.13	21.78
Chi-X				
Primary exchange	16.62	17.81	19.53	20.31
BATS	30.67	32.38	35.42	36.10
Turquoise	25.91	27.85	29.54	31.25
BATS				
Primary exchange	12.84	14.41	16.89	17.30
Chi-X	30.29	33.05	34.93	35.60
Turquoise	30.81	31.69	33.76	35.90
Turquoise				
Primary exchange	16.60	17.48	18.31	19.55
Chi-X	26.62	28.38	29.04	30.48
BATS	27.30	28.58	27.38	30.26

*Notes.* This table reports statistics on DCL measured as a percentage of the size of the trade that triggers the DCL measurement. Means of DCL are presented across all platforms (panel A), by platform (panel B; the platform under consideration being the quote venue ( $qv$ )), and by pair of platforms (panel C) for different time windows (10, 20, 50, and 100 ms). DCL is first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted DCL values are averaged for each stock over the entire month, and equally weighted means across stocks are calculated.

Table 3 has the same structure as Table 2 except that it reports figures based on DCL as a fraction of pretrade depth rather than trade sizes. It also includes figures on liquidity resupply after cancellations of duplicated orders (i.e., order book refills). The DCL mean percentages of panel A of Table 3 give a sense of the fraction of liquidity supplied to the consolidated market that is likely to disappear from one venue because of the execution of a duplicated limit order on another venue. According to those figures, the average level of DCL stands at 4.04% of the depth displayed in the consolidated market, a level that does not change much as one moves from a 10-ms DCL measurement window to a 100-ms window. Further, panel A of Table 3 shows that the average refill rate for all stocks combined is negative. This suggests that our DCL measure is not contaminated by cancellations because of members repricing orders in response to trades on other venues. In contrast, the relatively low but strictly positive value of the

opposite-side liquidity provision rate indicates that if any liquidity resupply follows DCL, it rather happens on the other side of the market, suggesting that DCL could be related to the practice of crossvenue market-making strategies.

Panel B of Table 3, which reports DCL by platform, shows that DCL is much smaller on primary exchanges than on the three alternative venues. When we break down DCL by pairs of venues in panel C of Table 3, we find that the proportion of limit order volume that is removed by the same member on another platform ranges from roughly 2% to almost 9%. The highest levels of DCL are observed for pairs of alternative venues where the quote venue is BATS and the trade venue is either Chi-X or Turquoise. As in panel A of Table 3, the average value of the refill rate is negative for almost all pairs of platforms, whereas the average opposite-side liquidity provision rates by pairs of platforms are all positive. Pairs of alternative venues, more

**Table 3.** Average Level of DCL as a Percentage of Quantities Available in the Book

	Liquidity resupply in the next 10 ms, %					
	10 ms, %	Same side (refill rate)	Opposite side	20 ms, %	50 ms, %	100 ms, %
Panel A: All platforms						
All platforms	4.04	-0.34	0.73	4.20	4.26	4.34
Panel B: By platform						
Primary exchange	3.43	-0.55	0.61	3.54	3.48	3.54
Chi-X	6.58	-0.78	1.58	7.06	7.47	7.61
Turquoise	5.98	0.58	0.82	6.22	6.54	6.72
BATS	6.92	-0.54	0.56	7.56	8.19	8.51
Panel C: By pair of platforms (quote venue and trade venue)						
Primary exchange						
Chi-X	3.74	-0.48	0.69	3.87	3.92	4.02
BATS	1.96	-0.19	0.37	2.00	1.69	1.50
Turquoise	3.30	-0.57	0.49	3.38	3.34	3.37
Chi-X						
Primary exchange	6.61	-0.86	1.31	7.11	7.58	7.80
BATS	5.25	-1.03	0.98	5.56	5.48	4.97
Turquoise	6.31	-0.31	1.90	6.51	6.63	6.60
BATS						
Primary exchange	6.19	-0.68	0.46	6.82	7.54	7.93
Chi-X	8.50	-1.41	1.00	9.39	9.77	9.72
Turquoise	8.55	-0.86	0.66	8.79	9.02	9.21
Turquoise						
Primary exchange	5.86	0.65	0.67	6.07	6.45	6.73
Chi-X	5.99	-0.33	1.27	6.28	6.34	6.30
BATS	4.94	-0.89	0.99	5.13	5.03	5.17

*Notes.* This table reports statistics on DCL measured as a percentage of quantities available in the order book of the quote venue prior to executions on the trade venue. Means of DCL are presented across all platforms (panel A), by platform (panel B; the platform under consideration being the quote venue ( $qv$ )), and by pair of platforms (panel C) for different time windows (10, 20, 50, and 100 ms). For DCL at the 10-ms horizon, the table also reports average rates of liquidity resupply within the next 10 ms (refill rates on the same side of the market and liquidity provision on the opposite side). Those rates are expressed as a percentage of DCL and winsorized at the 99% level. DCL and liquidity resupply rates are first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month, and equally weighted means across stocks are calculated.

specifically those involving Chi-X, exhibit the highest opposite-side liquidity provision rates. This again suggests that DCL plays a role in crossvenue market-making strategies, especially on alternative venues.

We have also computed the same measures as in Tables 2 and 3 but for a modified DCL measure. The DCL metric that we have worked with thus far (i.e., Equation (1)) subtracts the aggregate quantity traded in the interval from the difference between pre- and post-trade liquidity outstanding so as not to include involuntary reductions in liquidity associated with trades in the DCL measure. However, some of these trades may have been executions of genuine duplicated orders by counterparties with fast smart order routing technology (i.e., by agents whose technology is fast enough to allow them to hit duplicate orders on multiple venues before the liquidity suppliers can remove them). Thus, our DCL measure represents a lower bound on true duplicated liquidity. To provide an upper bound, we also compute summary statistics for a DCL measure that is just the change in liquidity pretrade to posttrade. This modified measure implicitly assumes that all executions against this member and in this stock in the interval were of duplicated limit orders. The adjustment roughly doubles the level of DCL measured as a fraction of outstanding liquidity from just over 4% (measured across all stocks) to almost 9%. On some markets and some venues, DCL reaches 15%. Thus, factoring volume executed on the quote venue into the DCL definition significantly increases the magnitude of DCL. Performing the same adjustment to our DCL measure based on trade size leads to statistics in which DCL rises from roughly 19% to 24%. Thus, there is an increase here too but proportionately less big.<sup>11</sup>

Returning to the original DCL measure, we proceed to investigate the variation of DCL with stocks' activity levels. Table 4 displays the average level of DCL by market value tercile, volatility tercile, and market fragmentation tercile. These levels are again expressed as percentages of both trade size and depth. For each scaling factor, Table 4 provides two columns: one reporting the average DCL observed on a single quote venue as in Equations (2) and (3) and another reporting average DCL aggregated across all possible quote venues.

Regarding our original DCL measures, which focus on a single quote venue at a time, differences in DCL expressed as a fraction of depth are in general not very large, but there is a tendency for DCL to rise with market capitalization. This tendency is much clearer when DCL is measured relative to trade size. Table 4 also demonstrates that DCL is negatively related to volatility in a stock, likely because when volatility is high, a market-making strategy that leaves multiple orders exposed on various venues is very costly. Finally, as one would expect, the bottom of Table 4 shows that DCL is larger on average for stocks with more

fragmented trading. Presumably, when volume is dispersed across venues, liquidity suppliers find it profitable to provide consistent liquidity across those venues through order duplication.

ESMA (2016) finds that the cross-stock covariances of order duplication intensity with market cap, volatility, and fragmentation have the same sign as the covariances between our DCL measure and those variables. They also find that the likelihood of duplicate orders being canceled also tends to rise with market cap and fragmentation. Thus, their results and our results are consistent.

In addition, to dig further, we also take an aggregate perspective and focus on the evolution of a member's *consolidated depth across all venues* around a trade. In particular, we study how a member's offering of market depth across all venues (i.e., all of their outstanding limit orders on the relevant side of the market across all venues) evolves in the time window before (i.e., at  $\tau$ ) to after (i.e., to  $\tau + 10$  ms) the trade taking place on a particular trade venue. We again scale this difference in depth either by a member's pre-event consolidated depth or by the size of the trade, and we control for trades against our member in the event window. Averages of this consolidated DCL are reported in the "In the consolidated book" columns in Table 4. We find that, on average, the liquidity withdrawn from all books immediately after a trade equals 6.62% of a member's consolidated depth and 59.09% of trade size. Because these numbers are larger than our single-venue DCL measures (4.04% and 18.89%, respectively), we find that members are not shifting limit orders from the quote venue to the trade venue. Instead, further liquidity is being withdrawn from all venues. The consolidated DCL as a fraction of trade size is, on average, three times greater than the single-venue DCL across all cases considered in Table 4. This indicates that DCL as a percentage of traded quantities is expected to increase with the number of venues available for order duplication.

Finally, we build on Section 4.2 and study whether orders that are canceled in the consolidated order book are refilled within the 10 ms following the time window over which DCL is measured (i.e., we analyze the "refill rate"). On average, we find a negative refill rate of  $-2.84\%$  of the globally canceled liquidity of that member, indicating that members continued canceling liquidity in the 10 ms after the trade interval.

#### 4.5. Alternative Explanation: Is DCL Really Canceled?

In this subsection, we discuss and rule out a possible alternative explanation for the magnitude of DCL. One possibility is that members move their orders from the quote venue to the "venue where the action takes place" (i.e., the trade venue) in order to increase their

**Table 4.** Average Level of DCL per Market Value Tercile, Volatility Tercile, and Fragmentation Tercile

Tercile	Range	Average DCL as a percentage of trade size (10 ms)		Average DCL as a percentage of pretrade liquidity (10 ms)	
		On the quote venue	In the consolidated book	On the quote venue	In the consolidated book
Market value tercile	Market value range (EUR mn)				
1	195–1,833	16.17	51.11	3.45	5.93
2	1,989–5,846	17.98	56.92	3.86	6.61
3	6,152–118,942	22.42	68.91	4.79	7.31
Volatility tercile	Daily volatility range, %				
1	0.0706–0.1253	22.74	58.03	4.96	6.40
2	0.1266–0.1549	18.78	51.60	3.97	6.02
3	0.1549–0.3266	15.04	67.36	3.17	7.42
Fragmentation tercile	Fragmentation index range				
1	1.0604–1.5520	7.18	40.18	1.68	4.88
2	1.5553–2.0663	15.10	50.59	3.35	6.14
3	2.0831–3.0714	33.90	85.61	7.00	8.78
All stocks		18.89	59.09	4.04	6.62

*Notes.* This table reports statistics on DCL by market value tercile, volatility tercile, and fragmentation tercile. DCL is here measured as a percentage of both the pretrade quantity posted by a member in the order book and as a percentage of the size of the trade that triggers the measurement of DCL. In each case, DCL is computed in two ways: first, by considering the cancellations on a single venue (the quote venue) as in Tables 2 and 3 and second, by aggregating cancellations across all venues where the member may have posted duplicate orders (consolidated book). DCL is first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members are constructed, where the weight for a member is equal to that member’s average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month, and equally weighted means across stocks are calculated.

execution probability. In that event, what we call duplicated-then-canceled liquidity would simply reflect a reshuffling of liquidity toward the trade venue. To study this alternative explanation, we first check whether orders canceled on the quote venue (DCL) are swiftly resubmitted on the trade venue in the same and next 10-ms windows. According to our observations, this is not the case. On average, across all stocks, 15.6% of the DCL measured on the quote venue is also canceled by the same member on the trade venue, and refill rates on the trade venue in the next 10 ms are close to zero.

## 5. Market Member Classification and DCL Level by Member Category

The ESMA data set contains all market members active on each trading venue during May 2013. There are 388 members in total for our 91 sample stocks. For each message in the data set, those market participants are identified by anonymized member identifications at several levels of granularity. First, each account for a particular member on a given venue is identified by a specific identification, which we call the unique identification. Second, all accounts of a given member on a given venue are identified with a common venue-specific identification designated as the account identification. Last, if a market participant is a member of several venues, all of the accounts of that member are

identified on all venues with a common crossvenue identification designated as the group identification. This group identification allows us to follow a market participant across venues. In addition, the data set provides information about member capacities. For each message, a flag indicates whether the member submitted the message as principal or agent.

From there, we establish and use three member classifications: (1) a slow/fast trader classification based on the HFT identification strategy established by ESMA, (2) a distinction between local members (that is, members acting on a single venue) and global members (that is, members trading across venues), and (3) a liquidity supplier/taker distinction.

### 5.1. Slow/Fast Trader Identification

According to MiFID II (cf. Article 4(1)(40)), an HFT technique is “an algorithmic trading technique characterized by: (a) infrastructure intended to minimize network and other types of latencies, including at least one of the following facilities for algorithmic order entry: colocation, proximity hosting or high-speed direct electronic access; (b) system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and (c) high message intraday rates which constitute orders, quotes or cancellations.” As HFT is a rather recent phenomenon, definitions are still evolving, and the academic literature adopts a variety of methods to classify

market participants as HFTs or non-HFTs; however, none of them are perfect.

Two main approaches are often used and sometimes combined. First, firms may be classified as either HFT or non-HFT based on public information about their primary business and the types of algorithms or services that they use. This approach will be referred to as the direct approach. Second, an analysis of firms' trading strategies (e.g., order placement and cancellation) can also allow a researcher to identify HFTs, and we refer to this as the indirect approach. HFT strategies are often characterized by a very short order lifetime (Hasbrouck and Saar 2013), a high order-to-trade ratio (Hendershott et al. 2011), and an inventory management policy that leads to traders carrying no significant positions overnight (Jovanovic and Menkveld 2016, Kirilenko et al. 2017). In the search for a more precise HFT classification, these criteria are sometimes combined. For example, Carrion (2013) and Brogaard et al. (2014) use a data set from Nasdaq, where Nasdaq defined a firm as an HFT both based on the quantitative properties of that firm's order submissions and trading behavior and based on the more general information on the firm's business model.

Our approach to categorizing firms by speed consists of two steps. First, we identify a set of fast traders using the indirect approach of Bouveret et al. (2014) based on the lifetime of orders.<sup>12</sup>

Bouveret et al. (2014) use the same data as we do, and they classify members as fast traders if the 10% quickest order modifications and cancellations in a given stock occur no more than 100 ms after the initial submission.<sup>13</sup> Such a criterion indicates that the member under consideration possesses fast trading technology, even if they do not use it at all times. We follow Bouveret et al. (2014) in choosing a fast trader identification based on the lifetime of orders because our main concern is trading speed, regardless of trading strategy. Criteria based on inventory management may identify fast traders implementing market-making strategies but not necessarily other fast traders. An identification based on order-to-trade ratios could also be biased, as slow traders with very few trades could be wrongly identified as fast. It is worth noting that Bouveret et al. (2014) find that just over 40% of value traded is done by fast traders using this approach.<sup>14</sup>

The fast trader flag is established by group identification, separately by capacity (agent or principal), and separately by stock. Therefore, a member may be a fast trader for some stocks and not for others, and for a given stock, a member may be defined as a fast trader when trading as principal but not when trading as agent. However, if a given market participant is flagged as a fast trader for their proprietary activity in stock  $i$  on venue  $v$ , they will be flagged the same way for proprietary activity on the other trading venues.

Second, we subdivide the population of fast traders into two distinct groups: HFTs and other algorithmic traders. This second step to the identification process, which is based on the direct approach of Bouveret et al. (2014), results in a list of 21 HFT firms. This list is built using firms' websites and the financial press to identify each firm's primary business, the use of services to minimize latency, and membership of the European Principal Trader Association. Any fast trading firm that is on this list and is trading as principal is defined as an HFT.

We then define algorithmic traders as the residual subset of fast traders who are not identified as HFTs. These firms are essentially investment banks. In common usage, algorithmic trading is any type of computer-based trading, including HFT. In our paper, for clarity, ATs and HFTs are two nonoverlapping groups of fast traders.

## 5.2. Global/Local Member Identification

Not all market participants are active on multiple venues during our sample period. Of the 388 participants, 307 trade on only one venue, whereas 81 trade on multiple platforms. The distinction between members trading at several locations, hereafter called global members, and members trading in a single market, hereafter referred to as local members, is instrumental to our study as DCL is defined as a consequence of multimarket trading strategies. We, therefore, classify global members as market participants who trade on at least two markets and execute more than 10% of their trading volume away from their main trading venue. Any member trading more than 90% of their volume in one market is classified as a local member. This classification is established by group identification, capacity, and stock.

## 5.3. Liquidity Supplier/Taker Identification

Duplicated limit orders are the outcome of trading strategies in which liquidity is offered at several locations in order to minimize nonexecution risk or equivalently, to capture fragmented market order flow. As such, DCL is expected to be mainly generated by traders implementing market-making strategies. For that reason, it seems relevant to distinguish members who are mainly passive in their trading strategies from those who are mainly active. The former will be referred to as liquidity suppliers, and the latter will be referred to as liquidity takers. A member is considered an LS (LT) if she is the passive (active) counterpart in more than 50% of her total consolidated trading volume when trading as principal. Finally, it is important to note that any member who is trading as an agent is always considered an LT as agents are executing position changes on behalf of clients rather than taking the other side of public orders and thus, seeing their own account affected. This classification is again established by member, by capacity, and on a stock-by-stock basis.

### 5.4. Member Combined Classification

A particular member in our data may engage in both principal and agency trading. Where a member in a given stock engages in both, these activities are separated in the data set via the previously mentioned capacity flag, resulting in distinct member/capacity pairings for that member and that stock. Although Bouveret et al. (2014) argue that the capacity flag cannot be used without difficulty to identify HFTs when using a direct approach and looking across stocks, the capacity flag can still be used for analysis at the stock level. The AT, HFT, global, and liquidity supplier flags are then assigned to each member/capacity pairing on a stock-by-stock basis. As a result, the classification applied to our 388 members produces 8,568 triplets of member  $\times$  capacity  $\times$  stock combinations. Further, for the sake of simplicity, in the remainder of the paper, when we use the term “member” or “trader,” we mean a member/capacity pairing.

The scheme described above generates 16 categories of traders. These are presented in Table 5 along with the number of member  $\times$  capacity  $\times$  stock combinations that fall into each category plus their market shares in trading. Note that there are 16 (not 24) categories as those trading as agents are never classified either as liquidity suppliers or as HFTs.

Based on the “No. of member/stock combinations” column in Table 5, the largest subgroups (3,259 of the total 8,568 combinations) correspond to slow local liquidity takers acting as agents, representing 38.04% of

the member  $\times$  capacity  $\times$  stock combinations. The slow local liquidity takers acting as principals account for 14.48% (1,241 combinations). Fast traders (ATs and HFTs) collectively represent 20.28% of the combinations. Global traders, including both slow traders and ATs/HFTs, account for 34.48% of the combinations. Lastly, liquidity suppliers, whether slow or fast, make up 18.83% of the total combinations, with 5.18% attributed to fast global liquidity suppliers, evenly split between ATs and HFTs.

In terms of trading volume (see the “Total” column under “Percentage in trading volume” in Table 5), ATs and HFTs account for 22.98% and 22.21% of the total traded value, respectively. These percentages are calculated by summing the values shown in the “Percentage in trading volume” column in Table 5 for the rows corresponding to ATs or HFTs. Trading volume from members acting as principals makes up 73.99% of the total volume, equally distributed between slow and fast traders. Global traders account for 72.81% of the total traded volume, whereas liquidity suppliers capture 25.47% of the total traded value distributed across the “slow,” “AT,” and “HFT” categories.

### 5.5. Market Fragmentation Level and Investor Clientele Segmentation

In terms of trading volume distribution, as reported in Table 5, 64.35% of the total volume is traded on primary exchanges, and Chi-X is the main alternative venue with a volume share of 20.91%. All platforms except

**Table 5.** Market Shares by Member Categories

Trading scope	Trading aggressiveness	Trading speed	Capacity	No. of member/stock combinations	Percentage in trading volume				
					Total	Primary exchange	BATS	Chi-X	Turquoise
Local trader	Liquidity taker	Slow	A	3,259	15.80	15.72	0.01	0.06	0.01
			P	1,241	4.88	4.31	0.02	0.37	0.18
		AT	A	247	3.79	3.78	0.00	0.01	0.00
			P	105	0.39	0.30	0.00	0.03	0.06
	Liquidity supplier	HFT	P	34	0.35	0.19	0.00	0.16	0.00
			Slow	P	545	0.99	0.81	0.01	0.10
		AT	P	122	0.50	0.36	0.01	0.12	0.02
			HFT	P	61	0.48	0.29	0.01	0.18
Global trader	Liquidity taker	Slow	A	527	3.23	1.87	0.24	0.89	0.22
			P	817	20.22	11.70	1.13	5.27	2.12
		AT	A	189	3.18	1.82	0.18	0.63	0.55
			P	231	7.37	4.19	0.42	1.59	1.18
			HFT	P	305	15.31	8.34	0.94	4.11
	Liquidity supplier	Slow	P	441	9.69	5.73	0.57	2.42	0.98
			AT	P	218	7.75	3.13	0.64	2.44
		HFT	P	226	6.06	1.81	0.76	2.54	0.94
			Total		8,568	100	64.35	4.92	20.91

*Notes.* This table displays the relative market size of each member category. Our member classification is established on a stock-by-stock basis and based on three criteria: local vs. global traders, liquidity suppliers vs. liquidity takers, and slow traders vs. ATs/HFTs. Flags for a given member on a given stock can also differ according to the member capacity (agent or principal). As a result, the “Number of member/stock combinations” column displays numbers of member  $\times$  capacity  $\times$  stock combinations. The right-hand side of the table reports the percentages of each category in total trading volumes with a breakdown by exchanges.

BATS have venue-specific local traders. Of our 388 market members, 307 trade on only one venue, with 297 trading only on the primary exchange, 8 trading only on Chi-X, and 2 trading only on Turquoise. Those 307 single-market players represent about 18% of total trading volume in our data set. Most of them typically trade only a few stocks, but 11 of the 307 are in the top 10% of market participants by activity. Those figures indicate that the market architecture is substantially fragmented, with more than a third of lit trading volumes executed outside primary exchanges and with single-venue trading occurring on every venue. However, this level of fragmentation is less than that observed in U.S. stock markets, where the share of primary exchanges is less than 40%. The fragmentation level in our data, which is intermediate between consolidation and total fragmentation, is referred to as market segmentation in Harris (1993).

In segmented markets, global traders in general and global liquidity suppliers in particular who, respectively, represent 72.81% and 23.50% of the total traded volume in our sample play a crucial role in partially reconsolidating the order flow and synchronizing prices. Among our 388 market members, 81 members trade on multiple platforms; 39 trade on all four platforms, 17 trade on three platforms only, and 25 trade on two platforms only. The 39 market participants trading on all venues account for about 71% of all trading volume. Twenty of the 39 are in the top 10% of market participants as measured by total trading activity.

The statistics in Table 5 also show that primary exchanges and alternative venues attract different trader clienteles. First, fast traders are much more active on alternative venues. The relative weights of ATs and HFTs are greater on BATS, Chi-X, and Turquoise, where their respective volume shares are 26.40% and 32.47% versus 21.09% and 16.53%, respectively, on primary exchanges. BATS and Chi-X have the highest shares of volume traded by HFTs (34.83% and 33.40%, respectively), whereas Turquoise has the highest share of volume traded by ATs (34.14%). Second, the weight of global traders is greater on alternative platforms, where they account for 96.02% of the volume. On the contrary, primary exchanges have the highest level of local trading volumes (25.77% versus 4.90% on Chi-X and 3.52% on Turquoise). Third, liquidity suppliers are more active on alternative venues, where they trade 37.45% of the volume.

In markets segmented in terms of both order flow and trader type, understanding to what extent and how duplicated limit orders are used in liquidity-providing strategies is of great interest. On the one hand, the swift cancellation of duplicates may adversely impact the execution costs of some traders, but on the other hand,

limit orders duplicated across venues may be a source of liquidity for the local traders of all venues.

### 5.6. DCL Level by Member Category

It is important to understand whether DCL is mainly because of some categories of members. Table 6 decomposes average DCL by members according to their trading scope (*local trader* and *global trader*) and trading aggressiveness (*liquidity taker* and *liquidity supplier*). We further distinguish according to their trading speed (*Slow*, *AT*, and *HFT*) and their capacity (*Agent* or *Principal*) in Table 6. The most interesting differences arise in two cases: (1) when comparing traders by speed and (2) when comparing members acting as principal with those acting as agents for their clients.

When comparing traders by speed, as we would expect, the average DCL for HFTs is (at 5.75% of their total pretrade depth) about 1.5 times the average DCL associated with ATs, which is, in turn, around 1.4 times the DCL from slow traders. Thus, HFT trading strategies involve greater duplicated liquidity. ESMA (2016) reports a similar finding for their direct analysis of order duplication.

When comparing members by trading capacity, DCL is typically higher when members are acting as principal rather than agent. Again, this supports the intuition that DCL activity is related to market-making strategies.

Let us recall that the starting point of a DCL calculation is a trade on a given venue. At the time of the trade, the passive counterparty may or may not have duplicated limit orders on the venue where DCL is measured. For that reason, we also provide, in Table 6, the percentage of trades for which there is order duplication on the quote venue. By definition, this percentage is extremely low for local traders (3.31%), but in those few cases where they duplicate orders, the average value of their DCL is more than half of that of global traders. Another striking case is that of members trading as agents. They duplicate limit orders far less often than members trading as principal (16.78% versus 51.23%), but when they do so, their level of DCL reaches one half of that of members trading as principal.

The fact that, on average, DCL differs systematically across member categories suggests that it may be important to control for such categories in our multivariate analysis.

## 6. Determinants of DCL

In this section, we set out to identify the determinants of DCL by testing the hypotheses developed in Section 2.2. As a reminder, those hypotheses posit that (Hypothesis 1) DCL increases with fragmentation; (Hypothesis 2) DCL increases in the relative tick size; (Hypothesis 3) frequent liquidity suppliers and (Hypothesis 4) traders acting as principal post more DCL; (Hypothesis 5) the



trade venue  $tv$  is  $PE$  and the venue on which we measure DCL (i.e., quote venue  $qv$ ) is  $ALT$ , and it is zero otherwise.  $ALTtoPE_{tv,qv}$  has a similar interpretation. The base case is where  $tv$  and  $qv$  are both  $ALT$ .

We further control for the DCL by other HFT members ( $DCL_{HFT\setminus i,t,-m}^{Others}$ ), other AT members ( $DCL_{AT\setminus i,t,-m}^{Others}$ ), and other slow traders ( $DCL_{Slow\setminus i,t,-m}^{Others}$ ), excluding member  $m$  (denoted by  $-m$ ) in period  $t$  for stock  $i$ . We also include a set of stock-time characteristics that are determinants of liquidity. Volatility  $\sigma_{i,t}$  is a price range computed as the difference between the highest and lowest prices of stock  $i$  over a 15-minute interval  $t$  scaled by the midpoint between the high and low values.  $VOLUME_{i,t}$  is the logarithm of the total euro volume traded in stock  $i$  on the four venues over period  $t$ .  $PRICE_{i,t}$  is the last crossvenue log midquote on the day of period  $t$  for stock  $i$ .

Finally, we control for past and contemporaneous order imbalance denoted by  $IMB_{i,t-1}$  and  $IMB_{i,t}$ , respectively, to make sure that DCL is not driven by trade-conveyed informational effects.  $IMB_{i,t}$  is the absolute value of the difference between aggressive buy and sell trading volumes expressed as a percentage of the total traded volume on all platforms for stock  $i$  in period  $t$ .

Columns 2–5 in Table 7 display the results for our empirical model that measures DCL as a percentage of the size of the triggering trade, employing different time windows ranging from  $\Delta\tau = 10 - 100$  ms. We employ a Tobit model as our dependent variable, and it has truncation at zero (i.e., in many instances, there is no withdrawal of duplicated liquidity (DCL = 0)). The last column in Table 7 presents the results where we scale DCL by pretrade liquidity. Here, we use a Tobit model with truncations at zero and one as the upper limit in this case is where all pretrade liquidity is

withdrawn (i.e., DCL = 1). For the model that uses DCL scaled by pretrade liquidity, we include trade size on the trade venue as an explanatory variable. We denote this variable by  $TRADESIZE_{i,t,m}$ , and it equals the average size of the trades executed on  $tv$  and triggering DCL on  $qv$  for member  $m$ , stock  $i$ , and period  $t$ . Size is measured as the log of the euro value of the trade.

We first examine the impact of member characteristics—our key variables of interest. Consistent with Hypothesis 6, all columns in Table 7 show that trades where limit orders posted by fast traders (both HFTs and ATs) are executed lead to significantly more DCL than otherwise similar trades against slow trader liquidity (the base case) and that HFTs post much more DCL than ATs, with a statistical significance at the 1% level.

Results in columns 2–5 in Table 7 show that HFTs on average withdraw 22 percentage points more of the size of the triggering trade compared with slow members (the reference case) and around 15 percentage points more when compared with AT members. ATs withdraw on average 5.5 percentage points more than slow traders. This is consistent with Hypothesis 6, and one interpretation of this is that fast traders are more active in making markets across venues than slow traders. Members acting as agent and liquidity suppliers withdraw 5 percentage points less and 8.5 percentage points more than principal traders and liquidity takers, respectively, consistent with Hypotheses 3 and 4. All of those effects are significant at the 1% level.

The last column in Table 7 shows that an HFT (AT) member withdraws 7.88 (2.80) percentage points more of their outstanding limit order quantity on venue  $qv$  following the execution of one of their limit orders on venue  $tv$  compared with a slow member in a similar situation. HFT members thus post just over five

**Table 6.** Average DCL as a Percentage of Pretrade Liquidity by Member Category

Trading characteristic	Member category	Average DCL as a percentage of pretrade liquidity (10 ms)	Percentage of cases with duplication	Average DCL as a percentage of trade size (10 ms)
Trading aggressiveness	Liquidity taker	3.69	34.42	13.53
	Liquidity supplier	3.81	54.84	18.43
Trading scope	Local	2.11	3.31	11.59
	Global	3.80	57.81	16.50
Trading speed	Slow	2.70	32.60	12.32
	AT	3.76	56.84	12.52
	HFT	5.75	53.65	16.87
Capacity	Agent	1.94	16.78	5.48
	Principal	3.93	51.23	17.56

*Notes.* This table reports statistics on DCL by member category. It also provides the proportion of trades for which pretrade liquidity is duplicated. DCL is first estimated for each member and each stock on a daily basis. Then, for each stock and each day, weighted averages across members in the considered category are constructed, where the weight for a member is equal to that member's average contribution to order book depth over the day. Finally, those daily weighted values are averaged for each stock over the entire month, and equally weighted means across stocks are calculated.

percentage points more DCL than AT members. In line with the results where DCL is expressed as a fraction of trade size, DCL as a percentage of pretrade liquidity is also more pronounced when a member (i) behaves as a liquidity supplier (2.58 percentage points) and (ii) acts as principal (2.03 percentage points; i.e.,  $AGENT = 0$ ).

The standardized absolute inventory variable has a significant and negative coefficient in our regressions. Hence, there is support for Hypothesis 5. The economic magnitude implied by the estimated coefficients is relatively modest, although with a one-standard-deviation increase in inventory leading to a fall in DCL of around 0.4 percentage points of the size of the triggering trade or 0.1 percentage points of pretrade liquidity. So, more extreme inventory positions are associated with a slightly smaller DCL. Thus, members do not use DCL to manage inventory in times when inventory is extreme; instead, they use DCL strategies when their inventories are relatively small. In other words, limit order traders are more likely to use DCL in the first phase of their market-making strategy (i.e., when they are happy to build trading positions) rather than in the second phase (when they are actively trying to unwind large positions). In that second phase, probably because of overexecution risk, it is likely that DCL is more costly.

We now turn to all other characteristics. The trade size estimate is only available for the regression where DCL is expressed as a fraction of pre-existing liquidity supply (the last column in Table 7), and it shows that larger trades are associated with greater DCL. Members have greater incentive to cancel resting orders when trade size on the trade venue is larger. Results show that when trade size doubles, DCL increases by 1.2 percentage points.

Table 7 also shows the results for the “platform characteristics.” Based on the second column in Table 7 ( $\Delta\tau = 10$  ms), the  $PEtoALT$  coefficient shows that DCL is less pronounced when the trade takes place on the primary exchange and the quote venue (where DCL is measured) is another venue compared with the base case  $ALTtoALT$ . The coefficient on  $ALTtoPE$  is significant, positive, and larger in magnitude than that on  $PEtoALT$  across all columns in Table 7. In sum, DCL is least pronounced when trades take place on the primary exchange and most pronounced for trades occurring on alternative venues and where the liquidity is then canceled on the primary exchange. This is in line with Hypothesis 8 and with crossplatform market making.

Our regression model controls for other member groups’ DCL activity on that day for that stock. In general, we find that a member’s DCL seems to comove with the DCL of other members. This effect is most pronounced when other ATs are active posters of DCL.

The significant positive coefficients on trading volume and fragmentation imply that DCL is greater for

stocks that are traded more heavily and on a dispersed set of platforms (in line with Hypothesis 1). Absolute order imbalance has a consistent and significant negative effect. We were concerned that the cancellation activity behind DCL might be generated by members revising stock valuations because of the information contained in trades. Neither past order imbalance nor contemporaneous order imbalance positively impact DCL, which is not in line with an information-based interpretation. There is weak evidence that DCL increases with an increase in the price range for stock  $i$  but much stronger evidence that DCL is smaller for stocks with larger tick sizes. The second result is inconsistent with Hypothesis 2, which suggested that DCL might be more intensively used when undercutting by price is more difficult.

Finally, there is a concave relationship between smart order routing and DCL. This generates small increases in DCL when smart order routing is scarce but rising, but it generates very large negative effects when smart order routing is large and rising (e.g., using the results from the last column in Table 7, if smart order routers made up 20% of the trade population, DCL as a fraction of liquidity would be four percentage points greater than if SOR was zero, whereas if SOR was at 60% of trading, DCL would be almost eight percentage points lower than if SOR was zero). So, when smart order routers are used extensively, we see low use of DCL, likely because of the multiple execution risk that SOR technology exposes the users of DCL to (in line with Hypothesis 7).

## 6.2. DCL Determinants by Member Category

Table 8 shows the results of Equation (6), where  $\Delta t = 10$  ms, for subsamples that focus on various member categories. This allows us to study whether specific explanatory variables are more important for some member categories. Column (1) in Table 8 focuses on all members who are “Fast traders only.” Columns (2) and (3) in Table 8 separate the fast traders into ATs and HFTs. Columns (4) and (5) in Table 8 display results for “Liquidity suppliers only” and “Fast liquidity suppliers only.”

The coefficient on  $HFT$  in column (1) in Table 8 shows that HFTs withdraw 2.6% more of the trade size on the quote venue than do ATs (i.e., the base case) following a trade on the trade venue. Liquidity suppliers exhibit substantially more DCL than other traders, in particular when they are HFTs. Compared with the second column in Table 7, which presents results for all member categories, some interesting differences in the coefficient signs and magnitudes can be observed. First, the positive coefficient on  $ALTtoPE$  in the regressions in Table 7 appears to be driven by the behavior of ATs (column (2) in Table 8 shows a positive coefficient of

**Table 7.** Tobit Regressions of DCL for Global Market Members

DCL measure Time window for DCL	$DCL^{tradesize}$				$DCL^{depth}$
	10 ms	20 ms	50 ms	100 ms	10 ms
Member characteristics					
<i>HFT</i>	0.2197*** (0.0040)	0.2266*** (0.0042)	0.2394*** (0.0044)	0.2066*** (0.0046)	0.0788*** (0.0014)
<i>AT</i>	0.0547*** (0.0039)	0.0536*** (0.0040)	0.0553*** (0.0043)	0.0534*** (0.0045)	0.0280*** (0.0014)
<i>Agent</i>	-0.0522*** (0.0072)	-0.0554*** (0.0075)	-0.0635*** (0.0079)	-0.0534*** (0.0083)	-0.0203*** (0.0025)
<i>Liquidity supplier</i>	0.0851*** (0.0034)	0.0937*** (0.0036)	0.0994*** (0.0038)	0.0935*** (0.0040)	0.0258*** (0.0012)
<i>Average inventory<sub>t-1</sub></i>	-0.0041*** (0.0029)	-0.0037*** (0.0030)	-0.0044*** (0.0032)	-0.0036*** (0.0033)	-0.0009*** (0.0010)
Trade characteristics					
<i>Trade size<sub>t</sub></i>					0.0119*** (0.0006)
Platform characteristics					
<i>PE to alternative</i>	-0.0611*** (0.0040)	-0.0626*** (0.0041)	-0.0624*** (0.0044)	-0.0623*** (0.0046)	-0.0183*** (0.0014)
<i>Alternative to PE</i>	0.1100*** (0.0039)	0.1142*** (0.0040)	0.1221*** (0.0043)	0.1138*** (0.0045)	0.0267*** (0.0014)
Other market member DCL					
<i>DCL(other, HFT)<sub>t</sub></i>	5.28E-06*** (0.0000)	5.01E-06*** (0.0000)	4.52E-06** (0.0000)	0.00E+00** (0.0000)	0.0472*** (0.0023)
<i>DCL(other, AT)<sub>t</sub></i>	3.48E-05*** (0.0000)	3.99E-05*** (0.0000)	2.17E-05*** (0.0000)	0.00E+00*** (0.0000)	0.0540*** (0.0021)
<i>DCL(other, Slow)<sub>t</sub></i>	1.39E-05*** (0.0000)	1.39E-05*** (0.0000)	9.84E-06*** (0.0000)	0.00E+00*** (0.0000)	6.40E-06 (0.0000)
Order flow characteristics					
<i>Volume<sub>t</sub></i>	0.0165*** (0.0023)	0.0174*** (0.0023)	0.0178*** (0.0025)	0.0162*** (0.0026)	0.0047*** (0.0008)
<i>Order imbalance<sub>t</sub></i>	-0.0383*** (0.0100)	-0.0403*** (0.0103)	-0.0428*** (0.0110)	-0.0393*** (0.0115)	-0.0092*** (0.0034)
<i>Order imbalance<sub>t-1</sub></i>	0.0009 (0.0094)	0.0001 (0.0098)	0.0006 (0.0103)	0.0001 (0.0109)	-0.0009 (0.0032)
<i>Fragmentation<sub>t-1</sub></i>	0.0060*** (0.0037)	0.0071*** (0.0038)	0.0072*** (0.0041)	0.0070*** (0.0043)	0.0020*** (0.0013)
<i>SOR<sub>t-1</sub></i>	1.2192*** (0.0850)	1.3062*** (0.0897)	1.3998*** (0.0955)	1.1862*** (0.1005)	0.3755*** (0.0275)
<i>(SOR<sub>t-1</sub>)<sup>2</sup></i>	-3.1961*** (0.4605)	-3.5307*** (0.4934)	-3.8082*** (0.5260)	-3.4107*** (0.5550)	-0.8473*** (0.1388)

0.83 for *ALTtoPE*) as the coefficient is largely negative and around the same magnitude as *PEtoALT* for HFTs (column (3) in Table 8). The negative coefficients of both *PEtoALT* and *ALTtoPE* in the case of HFTs (column (3) in Table 8) indicate that the level of DCL posted by HFTs is the highest when neither exchange involved in the duplication strategy is the primary market. Second, comovement of DCL is most

pronounced among own-member types, with  $DCL_{AT}^{Others}$  and  $DCL_{HFT}^{Others}$  being significant in columns (2) and (3) in Table 8, respectively.

Regarding order flow characteristics, trader inventory, and stock volatility, the coefficients of Table 8 are consistent in sign with those in Table 7. The coefficients on tick size are significantly negative for ATs and (fast) liquidity suppliers.

**Table 7.** (Continued)

DCL measure	$DCL^{trade\ size}$				$DCL^{depth}$
	10 ms	20 ms	50 ms	100 ms	10 ms
Stock characteristics					
<i>Price range</i> <sub><i>t</i>-1</sub>	0.0473* (0.1378)	0.0344 (0.1456)	0.0333 (0.1558)	0.0334 (0.1637)	0.0266*** (0.0499)
<i>Price</i>	0.0092 (0.0579)	0.0198 (0.0601)	0.0158 (0.0638)	0.0197* (0.0670)	-0.0053 (0.0201)
<i>Tick</i>	-128.66*** (86.0085)	-124.45*** (89.4588)	-123.66*** (94.9390)	-123.35*** (99.1710)	-15.1442*** (27.8036)
Fixed effects					
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
15-Minute period fixed effects	Yes	Yes	Yes	Yes	Yes
Pseudo- $R^2$ , %	8.35	7.94	7.52	7.38	9.69

*Notes.* This table reports the conditional marginal effects estimated from Tobit regressions of 15-minute DCL by member, stock, and pairs of platforms on various factors. DCL is computed in several ways: first, as a fraction of trade size over four different time intervals (10, 20, 50, and 100 ms) and second, as a fraction of pretrade depth on the quote venue at the 10-ms horizon. DCL is computed only using trades of global members. Each pair of platforms consists of the trade venue (i.e., the venue where the member was passively executed) and the quote venue (i.e., the venue where the member’s liquidity is potentially withdrawn). Reported coefficients are the marginal effects of the explanatory variables on DCL conditional on DCL being positive. The independent variables of interest comprise an HFT dummy equal to one for HFT members, an AT dummy equal to one for AT members, an agent dummy equal to one for a member trading as an agent, a liquidity-supplier dummy equal to one for members identified as liquidity suppliers, the member’s lagged average standardized absolute inventory, a *PE-to-Alternative* dummy equal to one when the trade venue is the primary exchange and the quote venue is an alternative platform, an *Alternative-to-PE* dummy equal to one when the trade venue is an alternative platform and the quote venue is the primary exchange, a lagged fragmentation index, the lagged value of an SOR proxy, and its squared value. The control variables include a lagged price range, the lagged and contemporaneous values of the imbalance between buy and sell orders as a percentage of the total traded volume, the log of the contemporaneous total traded volume, the log of the closing price, the relative tick size, the contemporaneous DCL measured for other HFT members, the contemporaneous DCL measured for other AT members, and the contemporaneous DCL measured for other slow traders. When DCL is measured as a fraction of pretrade quantities in the book of the quote venue, the average size of the trades triggering the DCL observation is also controlled for, and the Tobit specifications are double censored with a lower bound set to zero and an upper bound set to one. DCL as a percentage of trade size is winsorized at the 99% level. Standard errors are reported in parentheses.

\*Statistical significance at the 10% level; \*\*statistical significance at the 5% level; \*\*\*statistical significance at the 1% level.

## 7. Impact of the Cancellation of Duplicated Limit Orders on Trading Costs

Finally, we analyze how strategies involving the cancellation of duplicated orders affect the trading costs of liquidity takers. More specifically, we investigate the impact of DCL on effective spreads for various LT groups. We choose to focus our analysis on the DCL of HFTs because our preceding results imply that HFTs are the main drivers of DCL. On the one hand, we might expect markets where more duplicated orders are submitted to be those in which “genuine” liquidity is harder to measure, and so, execution cost management might be more difficult. This may result in additional costs of trading. But, on the other hand, given that DCL has been shown to contribute to more efficient crossvenue market making, it may help to reduce trading costs.

We test these hypotheses by running panel regressions of the effective spreads paid by LTs on the level of DCL submitted by HFTs in the previous period. The analysis is conducted on measures of spreads and DCL aggregated by 15-minute period, stock, and venue. For each 15-minute period and stock, we compute the

average DCL of HFTs for a particular venue by taking the measures computed earlier for pairs of trade venues and quote venues, fixing a particular quote venue, and aggregating across trade venues. The specification is as follows:

$$\begin{aligned}
 ES_{i,t,k} = & a + b \sigma_{i,t} + c VOLUME_{i,t,k} + d PRICE_{i,t} \\
 & + e TRADESIZE_{i,t,k} + f PE_{i,k} + g \hat{DCL}_{i,t-1,k}^{HFT,depth} \\
 & + h ES_{i,t-1,k} + \varepsilon_{i,t,k}.
 \end{aligned} \tag{8}$$

In this equation, the dependent variable  $ES_{i,t,k}$  is the average effective spread expressed as a percentage of the midquote for stock  $i$  in 15-minute period  $t$  on venue  $k$ . Our variable of interest is  $\hat{DCL}_{i,t-1,k}^{HFT,depth}$ , which represents the number of duplicated orders that are voluntarily canceled by HFTs for stock  $i$  on venue  $k$  over the previous 15-minute period. We take the DCL of the previous period with the intention of identifying causal effects of DCL on spreads. We may, however, still have an endogeneity problem as the decision to duplicate limit orders across venues in period  $t - 1$  might be determined by liquidity factors at play at both  $t - 1$  and  $t$ . To address this concern, we estimate a two-stage least squares regression model. In the first stage, we model

**Table 8.** Tobit Regressions of  $DCL^{tradesize}$  by Member Subsamples

	(1)	(2)	(3)	(4)	(5)
	Fast traders only	ATs only	HFTs only	Liquidity suppliers only	Fast liquidity suppliers only
Member characteristics					
<i>HFT</i>	2.6085*** (0.0867)			3.0612*** (0.094)	2.3394*** (0.1043)
<i>AT</i>				1.0864*** (0.091)	
<i>Agent</i>	-0.5097*** (0.2095)	-0.1170*** (0.1430)			
<i>Liquidity supplier</i>	1.7480*** (0.0909)	1.2700*** (0.0856)	2.1791*** (0.1645)		
<i>Average inventory</i> <sub><i>t</i>-1</sub>	-0.0847*** (0.0733)	-0.0546*** (0.0655)	-0.1056*** (0.1250)	-0.0401*** (0.0629)	-0.1078*** (0.0864)
Platform characteristics					
<i>PE to alternative</i>	-1.0315*** (0.1037)	-0.3109*** (0.0874)	-2.4181*** (0.1902)	-0.2572*** (0.0914)	-0.3579*** (0.1300)
<i>Alternative to PE</i>	-0.1231*** (0.1021)	0.8394*** (0.0862)	-3.0341*** (0.1863)	0.6400*** (0.0866)	-0.9881*** (0.1215)
Other market member DCL					
<i>DCL(</i> other, <i>HFT</i> ) <sub><i>t</i></sub>	0.0001** (0.0002)	-0.0002 (0.0010)	0.0004*** (0.0004)	0.0001* (0.0002)	0.0002*** (0.0003)
<i>DCL(</i> other, <i>AT</i> ) <sub><i>t</i></sub>	0.0007*** (0.0005)	0.0007*** (0.0004)	0.0002 (0.0010)	0.0004*** (0.0005)	0.0006*** (0.0006)
<i>DCL(</i> other, <i>Slow</i> ) <sub><i>t</i></sub>	0.0000 (0.0004)	0.0000 (0.0005)	0.0000 (0.0006)	0.0001*** (0.0003)	0.0000 (0.0004)
Order flow characteristics					
<i>Volume</i> <sub><i>t</i></sub>	0.4610*** (0.0587)	0.2877*** (0.0493)	0.5493*** (0.1069)	0.2839*** (0.0500)	0.3633*** (0.0693)
<i>Imbalance</i> <sub><i>t</i></sub>	-0.7223*** (0.2617)	-0.3034*** (0.2152)	-1.2870*** (0.4873)	-0.6422*** (0.2178)	-0.7402*** (0.3045)
<i>Imbalance</i> <sub><i>t</i>-1</sub>	-0.1642*** (0.2477)	-0.1223*** (0.2031)	-0.1578 (0.4633)	-0.1286*** (0.2057)	-0.2252*** (0.2880)
<i>Fragmentation</i> <sub><i>t</i>-1</sub>	0.2308*** (0.0963)	0.2010*** (0.0804)	0.1933*** (0.1756)	0.1235*** (0.0828)	0.1912*** (0.1151)
<i>SOR</i> <sub><i>t</i>-1</sub>	21.4296*** (2.3603)	14.5770*** (1.8987)	15.9104*** (4.5371)	20.8824*** (1.8534)	23.3098*** (2.9374)
( <i>SOR</i> <sub><i>t</i>-1</sub> ) <sup>2</sup>	-70.5661*** (13.6432)	-42.9978*** (10.5246)	-59.0154*** (27.2590)	-55.4877*** (9.8539)	-83.2212*** (17.4226)

the DCL of HFTs for stock  $i$  on venue  $k$  at period  $t$  in a first-stage regression specified as follows:

$$\begin{aligned}
 DCL_{i,t,k}^{HFT,depth} = & \alpha_0 + \alpha_1 DCL_{-i,t,k}^{HFT,depth} + \alpha_2 DCL_{i,t-1,k}^{HFT,depth} \\
 & + \beta_1 FRAG_{i,t-1} + \beta_2 TICK_{i,t} + \beta_3 SOR_{i,t-1} \\
 & + \beta_4 SOR_{i,t-1}^2 + \gamma_1 VOLUME_{i,t} + \gamma_2 \sigma_{i,t-1} \\
 & + \gamma_3 PRICE_{i,t} + \eta_{i,t,k}.
 \end{aligned} \tag{9}$$

Following Hasbrouck and Saar (2013) and Degryse et al. (2015), we use two instrumental variables in

Equation (9): (1) the DCL of HFTs of all stocks in the same size tercile, stock  $i$  excluded, on the same venue over the same period, which is denoted  $DCL_{-i,t,k}^{HFT,depth}$ , and (2) the lagged DCL of HFTs for stock  $i$  on the same venue. Other variables are factors that have been demonstrated to determine DCL in Section 6, namely order flow fragmentation, tick size, SOR activity, total trading volume, price range, and price level, which are measured by the variables  $FRAG$ ,  $TICK$ ,  $SOR$ ,  $VOLUME$ ,  $\sigma$ , and  $PRICE$ , respectively, in the same way as in Equation (6). We include stock fixed effects and intraday

**Table 8.** (Continued)

	(1)	(2)	(3)	(4)	(5)
	Fast traders only	ATs only	HFTs only	Liquidity suppliers only	Fast liquidity suppliers only
Stock characteristics					
<i>Price range</i> <sub><i>t</i>-1</sub>	1.9024** (3.3980)	1.1281* (3.1389)	2.8186* (5.5449)	1.5944** (3.1042)	2.0064** (4.0091)
<i>Price</i>	-0.0517 (1.4374)	-0.3779 (1.1921)	1.6374** (2.6526)	0.9892*** (1.3278)	0.2949 (1.7417)
<i>Tick</i>	-661.95 (1,934.42)	-921.15*** (1,736.65)	-383.75 (3,292.96)	-1,611.68*** (2,005.26)	-1,519.43*** (2,421.38)
Fixed effects					
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
15-Minute period fixed effects	Yes	Yes	Yes	Yes	Yes
Pseudo- <i>R</i> <sup>2</sup> , %	2.90	2.66	2.45	3.27	3.12

*Notes.* This table reports the conditional marginal effects estimated from Tobit regressions of 15 minute DCL by member, stock, and pairs of platforms on various factors. DCL is computed, at the 10-ms horizon, as a fraction of the size of the trade that triggered the DCL observation. DCL is computed only using trades of global members. The Tobit regressions are run for five different subsamples of members with double-censoring, the lower bound being set to 0 and the upper bound being set to 1. Each pair of platforms consists of the trade venue (i.e., the venue where the member was passively executed), and the quote venue (i.e., the venue where the member’s liquidity is potentially withdrawn). Reported coefficients are the marginal effects of the explanatory variables on DCL, conditional on DCL being positive. The independent variables of interest comprises an HFT dummy equal to one for HFT members; an AT dummy equal to one for AT members; an agent dummy equal to one for a member trading as agent; a liquidity-supplier dummy equal to one for members identified as liquidity suppliers; the member’s lagged average standardized absolute inventory, a *PE-to-Alternative* dummy equal to one when the trade venue is the primary exchange and the quote venue an alternative platform; an *Alternative-to-PE* dummy equal to one when the trade venue is an alternative platform and the quote venue is the primary exchange, a lagged fragmentation index, the lagged value of an SOR proxy, and its squared value. The control variables include a lagged price range; the lagged and the contemporaneous values of the imbalance between buy and sell orders as a percentage of the total traded volume; the log of the contemporaneous total traded volume; the log of the closing price; the relative tick size; the contemporaneous DCL measured for other HFT members; the contemporaneous DCL measured for other AT members; and the contemporaneous DCL measured for other slow traders. Standard errors are reported in parentheses.

\*Statistical significance at the 10% level; \*\*statistical significance at the 5% level; \*\*\*statistical significance at the 1% level.

time fixed effects, identifying each 15-minute period of the trading session. The values of  $DCL_{i,t,k}^{HFT,depth}$  predicted by Equation (9), denoted  $\hat{DCL}_{i,t,k}^{HFT,depth}$ , are then used with a lag as the main regressor in Equation (8). The first-stage regression shows highly significant estimates for both instrumental variables.<sup>17</sup> We believe that our approach, albeit imperfectly, considerably limits endogeneity issues and allows us to identify causal effects.

In Equation (8), we control for widely acknowledged determinants of spreads, such as volatility  $\sigma_{i,t}$ , measured as in Equation (6) as the 15-minute price range; log trading volume  $VOLUME_{i,t,k}$ , computed as the log of euro trading volume for stock  $i$  on venue  $k$  over period  $t$ ; and the same price level measure  $PRICE_{i,t}$  that we used in Equation (6). We also control for trade size and exchange type by introducing  $TRADESIZE_{i,t,k}$ , the average size of the trades that were used to construct the effective spread variable, and  $PE_{i,k}$ , a dummy equal to one when the venue for which we are computing effective spreads is the primary exchange and zero otherwise.

Furthermore, in order to examine whether the impact of the DCL of HFTs on effective spreads differs between

the primary exchange and alternative venues, we modify Equation (8) by interacting  $\hat{DCL}_{i,t-1,k}^{HFT,depth}$  with exchange dummies in the following way:

$$\begin{aligned}
 ES_{i,t,k} = & a + b\sigma_{i,t} + cVOLUME_{i,t,k} + dPRICE_{i,t} \\
 & + eTRADESIZE_{i,t,k} + fPE_{i,k} + g_1\hat{DCL}_{i,t-1,k}^{HFT,depth} \\
 & \times PE_{i,k} + g_2\hat{DCL}_{i,t-1,k}^{HFT,depth} \times ALT_{i,k} + hES_{i,t-1,k} \\
 & + \varepsilon_{i,t,k},
 \end{aligned} \tag{10}$$

where dummy  $ALT_{i,k}$  equals one when venue  $k$ , where effective spreads are measured, is an alternative exchange and zero otherwise. In other words,  $ALT_{i,k}$  simply equals  $1 - PE_{i,k}$ .  $g_1$ , the coefficient of the product of the DCL measure and the primary exchange dummy, identifies the impact of the DCL of HFTs on the effective spreads of the primary exchange, whereas  $g_2$ , the coefficient of the product of  $DCL$  and  $ALT$ , identifies the impact on the spreads of alternative exchanges.

In both Equations (8) and (10), autocorrelation is accounted for by including the first lag of the dependent variable, and intraday time fixed effects identifying each 15-minute period of the trading session are included. We run three versions of those two

regressions by computing average effective spreads for three groups of LTs differentiated by trading speed. Those three groups are slow liquidity takers, algorithmic liquidity takers, and HFT liquidity takers, respectively.

Table 9 contains estimates of our second-stage regressions (8) and (10). There are several familiar results in Table 9 (e.g., spreads decrease with volume and price level, increase with volatility and trade size, and are positively autocorrelated). We also see that they are, on average, larger on the primary exchange.

As for the coefficients on DCL, we find that when looking at all venues combined, greater DCL from HFTs leads to larger effective spreads for slow LTs and smaller effective spreads for fast LTs at a 1% level of statistical significance. The adverse effect on the trading costs of slow LTs is most severe on the alternative venues and is smaller on the primary exchange, where its economic significance is greatly reduced, and its statistical significance is at the 10% level only. Presumably, the DCL of HFTs is less detrimental to the effective spreads of slow traders on the primary exchange

because slow traders themselves form a greater part of the trader population there, and thus, the DCL of a subgroup of traders has smaller impact. Further, slow traders are typically less active on alternative exchanges. We should, therefore, consider the  $g_1$  coefficient as more relevant than the  $g_2$  coefficient for this group of traders. In this regard, the adverse effect of HFTs' DCL on the trading costs of slow LTs can be seen as small in both economic size and statistical significance.

Regarding the magnitude of the beneficial effect for fast traders, the economic impact, all venues considered, is similar for ATs and HFTs, with the  $g$  coefficient having similar values for both trader categories. However, when differentiating between the primary exchange and alternative venues, the picture changes. Whereas HFTs benefit from DCL everywhere, be it on the primary exchange or on alternative venues, ATs benefit from it on alternative exchanges only. On the contrary, their trading costs increase with the DCL of HFTs on the primary exchange. Thus, taken together, ATs may benefit or not depending upon their activities across primary and alternative exchanges. Weighing up algorithmic

**Table 9.** Impact of the  $DCL^{depth}$  of HFTs on the Effective Spreads of Liquidity Takers

	Effective spreads of slow LTs		Effective spreads of ATs/LTs		Effective spreads of HFTs/LTs	
	Eq. (8)	Eq. (10)	Eq. (8)	Eq. (10)	Eq. (8)	Eq. (10)
$Price\ range_t$	0.0694*** (0.0121)	0.0692*** (0.0121)	0.0494*** (0.0132)	0.0500*** (0.0132)	0.0643*** (0.3417)	0.0642*** (0.0078)
$Volume_t$	-0.0034*** (7.23E-05)	-0.0034*** (7.24E-05)	-0.0032*** (8.85E-05)	-0.0032*** (8.85E-05)	-0.0028*** (5.17E-05)	-0.0028*** (5.17E-05)
$Price$	-0.0068*** (9.18E-05)	-0.0068*** (9.18E-05)	-0.0072*** (0.0001)	-0.0072*** (0.0001)	-0.0006*** (6.98E-05)	-0.0006*** (6.98E-05)
$Trade\ size_t$	0.0011*** (8.60E-05)	0.0011*** (8.60E-05)	0.0007*** (9.78E-05)	0.0007*** (9.80E-05)	-0.0005*** (6.38E-05)	-0.0005*** (6.38E-05)
$PE$	0.0063*** (0.0003)	0.0077*** (0.0004)	0.0022*** (0.0003)	-0.0006 (0.0005)	0.0050*** (0.0002)	0.0061*** (0.0003)
$\hat{DCL}(HFT)_{t-1}$	0.0111*** (0.0012)		-0.0082*** (0.0015)		-0.0087*** (0.0008)	
$\hat{DCL}(HFT)_{t-1} \times PE$		0.0036* (0.0021)		0.0052** (0.0026)		-0.0141*** (0.0015)
$\hat{DCL}(HFT)_{t-1} \times Alternative$		0.0137*** (0.0014)		-0.0131*** (0.0017)		-0.0070*** (0.0009)
$Effective\ spread_{t-1}$	0.6992*** (0.0022)	0.6987*** (0.0022)	0.5696*** (0.0030)	0.5690*** (0.0030)	0.3140*** (0.0034)	0.3139*** (0.0034)
Fixed effects						
15-Minute period fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted $R^2$ , %	64.32	64.23	45.79	45.82	16.15	16.17

*Notes.* This table reports the second-stage OLS (Ordinary Least Squares) regression results of effective spreads by stock-venue combination on the DCL of HFTs by 15-minute periods for alternatively slow liquidity takers, algorithmic liquidity takers, and HFT liquidity takers. The DCL considered is the DCL of HFTs as predicted in a first-stage regression in which the instruments are the contemporaneous DCL of HFTs in other stocks of the same size tercile and the DCL of HFTs for the same stock in the previous 15-minute period. Dummies "PE" and "Alternative" indicate whether the observation stems from, respectively, the primary exchange or an alternative venue. The control variables include price range, traded volume, closing price, average trade size, and the lagged effective spread for that stock and trader group. Standard errors are reported in parentheses.

\*Statistical significance at the 10% level; \*\*statistical significance at the 5% level; \*\*\*statistical significance at the 1% level.

LTs' costs and benefits, the economic significance of the benefits on alternative exchanges is still larger than that of the cost increase on primary exchanges.

All in all, these findings in conjunction with those of Sections 5 and 6 suggest that DCL makes market-making strategies more effective on the exchanges where HFTs are the most active relative to other traders (i.e., the alternative venues). Those more effective market-making strategies benefit all fast traders on alternative exchanges, whereas they only benefit HFT liquidity takers on primary exchanges. The DCL of HFTs on primary exchanges adversely impacts the trading costs of both slow and algorithmic LTs, with ATs being more significantly affected than slow LTs. Considering that slow traders mainly trade on primary exchanges and that the significance of the adverse effect of DCL on their trading costs is weak there, we can conclude that the adverse effect of the cancellation of duplicated orders by HFTs is somewhat limited for slow LTs. In contrast, the fact that the cancellation of duplicated limit orders by HFTs on primary exchanges is more harmful for the trading costs of algorithmic LTs than for those of slow traders comes, to some extent, at a surprise. It opens questions for future research on the differences between algorithmic and HFT trading strategies and their interactions on primary exchanges.

## 8. Conclusion

The objective of this paper is to assess the scale of duplicated-then-canceled liquidity on equity markets and to examine its determinants. Our DCL measure is related to limit order duplication across venues. We define it to exist when in response to the execution of a limit order on a particular venue, the submitter of that order swiftly cancels similar limit orders on other venues. Such liquidity provision strategies are natural and valuable in a fragmented trading landscape, where some liquidity suppliers are market making across venues. Thus, on the one hand, duplicating orders across venues may benefit crossmarket liquidity by improving execution probabilities and by bringing liquidity to the single-market players of all venues, but on the other hand, DCL may lead some market participants to overestimate the true liquidity available in the marketplace.

By drawing on a unique data set that covers 91 European stocks listed in nine different countries and their trading on their respective primary exchanges plus the three main alternative trading venues in Europe (i.e., Chi-X, BATS, and Turquoise), we find that DCL is an economically significant phenomenon that deserves attention from market participants and regulators. In the presence of duplicated limit orders, for 100 shares traded on one venue, the submitter of the passive order removes on average around 19 shares from the order book of each competing venue.

In the cross-section, DCL is greater for larger, more fragmented stocks and for less volatile stocks. Further, DCL increases with trading volumes and trade size. It decreases when smart order routing is particularly prevalent. HFTs, traders acting as principals, and traders implementing multivenue market-making strategies post more DCL than others. Further, regarding HFTs, their use of DCL is the highest when they duplicate limit orders across alternative platforms. The preceding results are robust to changes in the time window used to measure DCL, and they are not significantly impacted by cancellations because of information-based quote updating in response to trades. These results together suggest that DCL is generated, in part at least, by liquidity-supplying traders (acting as principal) implementing market-making strategies to serve disconnected liquidity pools (e.g., when fragmentation is large); that HFTs contribute substantially to this activity; and that when market conditions do not favor duplication (e.g., when volatility is high), the use of the strategy declines. Further, when the liquidity demand side of the market is using tools that allow them to access all venues simultaneously (e.g., smart order routing), DCL reduces as it becomes risky and potentially costly for liquidity suppliers. More specifically, we provide evidence of a negative relation between a member's DCL and its stock inventory. This suggests that DCL is used to build inventory rather than to rebalance extreme inventories.

At the market level, over 4% of the consolidated depth is DCL, with this average percentage being greater on alternative venues (between 6% and 7%) than on primary exchanges (3.43%). Those figures are unlikely to be large enough either to counterbalance the depth improvement related to fragmentation reported in previous literature (e.g., Foucault and Menkveld 2008, Degryse et al. 2015, Gresse 2017) or to create significant "noise" in total liquidity measures. Furthermore, DCL does not necessarily affect all traders in the same way as fast traders using properly calibrated smart order routers may catch DCL before it is withdrawn.

In our final piece of analysis, we find that although DCL is associated with slow traders suffering marginally increased trading costs, at times when DCL is high, the execution costs of fast liquidity takers (i.e., ATs and HFTs) on alternative trading venues are significantly smaller. We interpret the latter as evidence that DCL contributes to the efficiency of market-making strategies that HFTs and others run on alternative exchanges, allowing counterparties access to better prices on average.

Overall, we show that duplicated-then-canceled liquidity is a significant phenomenon in European equity markets. A downside of its use is that naive measurement of consolidated liquidity may overestimate true liquidity in fragmented electronic markets. However, our final results suggest that it is an important tool in facilitating crossvenue market making, leading to

improved realized liquidity on alternative venues and greater true consolidated liquidity.

### Acknowledgments

The authors thank two anonymous reviewers, an anonymous associate editor, Panos Anagnostidis, Carlos Aparicio Roqueiro, Jos van Bommel, Antoine Bouveret, Agostino Capponi, Carole Comerton-Forde, Sarah Draus, Thierry Foucault, Monika Gehde-Trapp, Cyrille Guillaumie, Björn Hägströmer, Frank Hatheway, Peter Hoffmann, Charles Jones, Olga Klein, Christophe Majois, Katya Malinova, Giang Nguyen, Carol Osler, Christine Parlour, Tavy Ronen, Gideon Saar, Anne-Laure Samson, Duane Seppi, Andriy Shkilko, Elvira Sojli, Ingrid Werner, Christian Winkler, and Gunther Wuyts; participants of the Group of Economic Advisors at ESMA (European Securities and Markets Authority); seminar participants at Katholieke Universiteit Leuven; members of the Autorité des Marchés Financiers scientific board; and participants at the 35th Spring Association Française de Finance Conference, the 4th Women in Microstructure Meeting, the 2nd European Capital Market Workshop, the 2nd Sustainable Architecture for Finance in Europe Market Microstructure Conference, the 2018 Financial Management Association Annual Meeting, the 46th European Finance Association Meeting (Lisbon), and the 2020 Plato Market Innovator Conference for their comments. The authors also thank Yujuan Zhang for excellent research assistance and ESMA for providing access to the data. The views presented in this paper are those of the authors and do not necessarily reflect the views of ESMA.

### Endnotes

<sup>1</sup> Chen and Duffie (2021) show that market fragmentation leads to higher price efficiency as liquidity takers use SORs to tap liquidity at multiple venues at the same time. This highlights the limits of canceling at one venue following a trade at another venue.

<sup>2</sup> The primary exchanges are Euronext Amsterdam, Euronext Brussels, Euronext Lisbon, Euronext Paris, Deutsche Börse, Borsa Italiana, the London Stock Exchange, the Irish Stock Exchange, and the Spanish Stock Exchange.

<sup>3</sup> At the current date, Chi-X and BATS, renamed CXE and BXE, respectively, are operated by CBOE Global Markets as a result of CBOE taking over Bats Global Markets in 2017.

<sup>4</sup> Some of the market operators running the trading venues considered in our study also run dark pools, but the trading activity of those dark pools is not included in our data set.

<sup>5</sup> Either those stocks were not traded over several days, or they were not traded outside the primary exchange.

<sup>6</sup> This approach is used in the report by ESMA (2016), which uses the same data as we do to show that around 20% of all limit orders are duplicated, with the duplication strategy used more frequently by HFTs and for large cap stocks. They also show that following around 25% of all trades, the liquidity supplier cancels duplicate orders on other trading venues. We contributed to the development of the measures used in this ESMA report as independent experts (ESMA 2016).

<sup>7</sup> Other time intervals considered are 20, 50, and 100 ms. There are all below human reaction time.

<sup>8</sup> As market makers' posted price schedules are supposed to be regret free, such information-induced cancellations would be caused by trades against other members than the one for which we measure DCL.

<sup>9</sup> Order submissions are only counted toward the refill quantity if they are submitted within a certain distance of the midquote. This

distance is the same as that defined above for the DCL computation, and the midquote that we use is that observed at the end of the DCL measurement window.

<sup>10</sup> For the statistics in Table 2, aggregated DCL is scaled by the trading activity corresponding to all the trades that triggered a DCL computation for a given member on a given trade venue on a given day. For simplicity, this trading activity is referred to as trade size, and it should not be confused with the total trading volume of a given day on a given platform.

<sup>11</sup> Tables of summary statistics for the adjusted DCL measures that identical in structure to Tables 2 and 3 are available on request from the authors.

<sup>12</sup> We contributed to the preparation of this report as independent experts.

<sup>13</sup> One hundred milliseconds is clearly below human reaction time. For purposes of comparison, the average duration for a single blink of a human eye is 0.1–0.4 seconds or 100–400 ms according to the Harvard Database of Useful Biological Numbers.

<sup>14</sup> They also do some robustness checks, varying the 100-ms threshold, and they show that although fast trading intensity and the threshold are obviously positively related, the slope of the relationship is fairly flat between 50 and 250 ms.

<sup>15</sup> This type of measure is commonly used in the literature on market fragmentation (see Degryse et al. 2015 and Gresse 2017). In terms of interpretation, our FRAG index ranges from one to four, with one indicating no fragmentation or in other words, a consolidation of volumes on a single venue and four indicating maximum fragmentation: that is, volumes equally distributed across the four venues. An FRAG index of two would mean that the level of fragmentation is equivalent to the maximum level of fragmentation between two markets (i.e., 50% of the volumes on each).

<sup>16</sup> This is very similar to what other authors have done in the microstructure literature. For example, in their analysis of how low-latency activity affects market quality, Hasbrouck and Saar (2013) use the average low-latency trading in other stocks during the same time period as an instrument for low-latency trading in a given stock. Very similar approaches were also used by Degryse et al. (2015) and Buti et al. (2022).

<sup>17</sup> First-stage results are not reported here for the sake of brevity, but they are available on request.

### References

- Baldauf M, Mollner J (2021) Trading in fragmented markets. *J. Financial Quant. Anal.* 56(1):93–121.
- Baruch S, Glosten LR (2013) Flickering quotes. Working paper, Columbia Business School, Columbia, NY.
- Blocher J, Cooper RA, Seddon J, Van Vliet B (2016) Phantom liquidity and high frequency quoting. *J. Trading* 11(3):6–15.
- Bouveret A, Guillaumie C, Aparicio Roqueiro C, Winkler C, Nauhaus S (2014) Economic report: High-frequency trading activity in EU equity markets. ESMA Economic Report No. 2, ESMA, Paris.
- Brogaard J, Hendershott TJ, Riordan R (2014) High-frequency trading and price discovery. *Rev. Financial Stud.* 27(8):2267–2306.
- Buti S, Rindi B, Werner IM (2022) Diving into dark pools. *Financial Management* 51(4):961–994.
- Carrion A (2013) Very fast money: High-frequency trading on the NASDAQ. *J. Financial Markets* 16(4):680–711.
- Chen D, Duffie D (2021) Market fragmentation. *Amer. Econom. Rev.* 111(7):2247–2274.
- Dahlström P, Hägströmer B, Nordén LL (2024) The determinants of limit order cancellations. *Financial Rev.* 59(1):181–201.
- Degryse H, de Jong F, van Kervel V (2015) The impact of dark and visible fragmentation on market quality. *Rev. Finance* 19(4):1587–1622.

- ESMA (2016) Order duplication and liquidity measurement in EU equity markets. ESMA Economic Report No. 1, ESMA, Paris.
- Foucault T, Menkveld A (2008) Competition for order flow and smart order routing systems. *J. Finance* 63(1):119–158.
- Gresse C (2014) Market fragmentation and market quality: The European experience. Dufrénot G, Jawadi F, Louhichi W, eds. *Market Microstructure and Nonlinear Dynamics: Keeping Financial Crisis in Context* (Springer, Cham, Switzerland), 1–24.
- Gresse C (2017) Effects of lit and dark market fragmentation on liquidity. *J. Financial Markets* 35:1–20.
- Hansch O, Naik NY, Viswanathan S (1998) Do inventories matter for dealership markets? Evidence from the London Stock Exchange. *J. Finance* 53(5):1623–1656.
- Harris L (1993) Consolidation, fragmentation, segmentation and regulation. *Financial Markets Institutions Instruments* 2(5):1–28.
- Hasbrouck J, Saar G (2013) Low-latency trading. *J. Financial Markets* 16(4):646–679.
- Hendershott TJ, Jones CJ, Menkveld AJ (2011) Does algorithmic trading improve liquidity? *J. Finance* 66(1):1–33.
- Jovanovic B, Menkveld AJ (2016) Middlemen in limit order markets. Preprint, submitted June 21, <https://doi.org/10.2139/ssrn.1624329>.
- Kirilenko AA, Kyle AS, Samadi M, Tuzun T (2017) The flash crash: High frequency trading in an electronic market. *J. Finance* 72(3):967–998.
- Malinova K, Park A (2020) “Sniping” in fragmented markets. Working paper, University of Toronto, Toronto.
- Menkveld AJ (2013) High frequency trading and the new market-makers. *J. Financial Markets* 16(4):712–740.
- O’Hara M, Ye M (2011) Is fragmentation harming market quality? *J. Financial Econom.* 100(3):459–474.
- van Kervel V (2015) Competition for order flow with fast and slow traders. *Rev. Financial Stud.* 28(7):2094–2127.
- Yueshen BZ (2025) Queuing uncertainty of limit orders. *Management Sci.*, ePub ahead of print September 17, <https://doi.org/10.1287/mnsc.2023.03371>.