American Economic Review 2021, 111(7): 2247–2274
https://doi.org/10.1257/aer.20200829

Market Fragmentation[†]

By Daniel Chen and Darrell Duffie*

We model a simple market setting in which fragmentation of trade of the same asset across multiple exchanges improves allocative efficiency. Fragmentation reduces the inhibiting effect of price-impact avoidance on order submission. Although fragmentation reduces market depth on each exchange, it also isolates cross-exchange price impacts, leading to more aggressive overall order submission and better rebalancing of unwanted positions across traders. Fragmentation also has implications for the extent to which prices reveal traders' private information. While a given exchange price is less informative in more fragmented markets, all exchange prices taken together are more informative. (*JEL* D47, D82, G14)

In modern financial markets, many financial instruments trade simultaneously on multiple exchanges (Budish, Lee, and Shim 2019; Gresse 2012; Pagnotta and Philippon 2018). This fragmentation of trade across venues raises concerns over market depth. One might therefore anticipate that fragmentation worsens allocative efficiency through the strategic avoidance of price impact, which inhibits beneficial gains from trade (Vayanos 1999, Du and Zhu 2017). Fragmentation might seemingly, therefore, lead to less aggressive trade, which could in turn impair the informativeness of prices, relative to a centralized market in which all trade flows are consolidated. Perhaps surprisingly, we offer a simple model of how fragmentation of trade across multiple exchanges, despite reducing market depth, actually *improves* allocative efficiency and price informativeness.

In the equilibrium of our market setting, the option to split orders across different exchanges reduces the inhibiting effect of price-impact avoidance on total order submission. Though market depth on each exchange decreases with fragmentation, the common practice of order splitting allows traders to shield orders submitted to a given exchange from the price impact of orders submitted to other exchanges. This effect is sufficiently strong that fragmentation increases overall order aggressiveness.

*Chen: Graduate School of Business, Stanford University (email: dtchen@stanford.edu); Duffie: Graduate School of Business, Stanford University (email: duffie@stanford.edu). Mikhail Golosov was the coeditor for this article. We are grateful for research assistance from David Yang and for conversations with and comments from Mohammad Akbarpour, Bob Anderson, Markus Baldauf, Anirudha Balasubramanian, Jonathan Berk, Eric Budish, Peter DeMarzo, Joe Hall, Elvis Jarnecic, Charles-Albert Lehalle, An Qi Liu, Ananth Madhavan, Semyon Malamud, Albert Menkveld, Michael Ostrovsky, Mathieu Rosenbaum, Marzena Rostek, Satchit Sagade, Marcos Salgado, Yuliy Sannikov, Amit Seru, Yazid Sharaiha, Andy Skrzypacz, Bob Wilson, Milena Wittwer, and Haoxiang Zhu. Three anonymous referees provided detailed helpful comments and suggestions.

[†]Go to https://doi.org/10.1257/aer.20200829 to visit the article page for additional materials and author disclosure statements.

This in turn leads to a more efficient redistribution of unwanted positions across traders and causes prices, collectively across all exchanges, to better reflect traders' private information. Once fragmentation is sufficiently severe, however, any additional fragmentation can cause trade to become *too* aggressive, from the perspective of allocative efficiency. However, at least in the simple one-period version of our model, any degree of fragmentation is welfare-superior to a centralized market.

Our model abstracts from some important aspects of functioning financial markets. In particular, we do not consider the impact of fragmentation on exchange competition or transaction fees.[1] We also ignore the adverse impact of sniping by fast traders (Budish, Cramton, and Shim 2015; Malinova and Park 2019; Pagnotta and Philippon 2018). Given these and other limitations of our model, we avoid taking a policy stance on fragmentation. Our primary marginal contribution is to identify a potentially important distinct economic channel for the welfare implications of market fragmentation.

We now briefly summarize our model and the main results. A single asset is traded by N strategic traders participating on E exchanges. Before each round of trade, each strategic trader has a quantity of the asset which is privately observed. Each trader submits a package of limit orders (forming a demand function) to each of the exchanges, simultaneously. As in common practice (Wittwer 2021), orders to a given exchange cannot be made contingent on clearing prices at other exchanges. The objective of each strategic trader, given the conjectured order submission strategies of the other traders, is to maximize the total expected discounted cash compensation received for executed orders, net of the present value of asset holding costs that are quadratic in the trader's asset position, as in the one-exchange model of Du and Zhu (2017).

At each exchange, "liquidity traders" submit nondiscretionary market orders. The aggregate quantities of market orders submitted by liquidity traders to the various exchanges are exogenous random variables, independently and identically distributed across exchanges and periods. In a one-period setting, we also consider a version of the model with no liquidity traders, and a version in which liquidity traders who are local to each exchange are strategic with respect to order quantities. In any version of the model, because agents' preferences are quasilinear in cash and because total cash payments net to zero by market clearing, an unambiguous measure of allocative efficiency is the expected discounted sum of strategic traders' asset holding costs.

Price impact is increased by market fragmentation because of cross-exchange price inference, by which traders choose order submissions in light of the positive equilibrium correlation between exchange prices. For example, conditional on a clearing price on a given exchange that is lower than expected, a buyer expects to be assigned higher quantities on all exchanges. This effect dampens the aggressiveness of order submissions, which reduces market depth and heightens market impact, relative to a single-exchange setting. Despite this reduction in market depth, the ability to split orders across exchanges ensures that, in equilibrium, each strategic trader's overall order submission is more aggressive, resulting in a more efficient

[1] As shown by Budish, Lee, and Shim (2019), transaction fees are economically small.

allocation.[2] This natural implication of fragmentation is novel to this paper, as far as we know.

We solve both static and dynamic versions of the model. In the static model, as the number of exchanges increases, the equilibrium allocation becomes more efficient until a point at which trade becomes "too aggressive." We find that the socially optimal number of exchanges depends only on (i) the number of strategic traders and (ii) the ratio of the variance of the endowments of strategic traders to the variance of liquidity trade. We show that when there are more exchanges, the price on any individual exchange is less informative of the aggregate asset inventory of strategic traders, the key "state variable" of our model, yet the exchange prices taken together are more informative. Although allocative efficiency is maximal for a finite number of exchanges, price informativeness always improves with the number of exchanges.

In the dynamic version of the model, we show that market fragmentation still allows efficient trade, despite the associated cross-period cross-exchange price impact and despite within-period price impact that is even higher than in the static model. We do not solve for an equilibrium of the dynamic model for an arbitrary number E of exchanges, given the difficult-to-solve infinite regress of beliefs about beliefs concerning the aggregate asset inventory of strategic traders. Rather than addressing equilibria for general E, we instead construct an equilibrium for a specific number E of exchanges with the property that the associated equilibrium is perfect Bayesian and implements efficient trade. This equilibrium is tractable because efficient trade dramatically simplifies the inference problem of each trader, given that the sum of exchange prices perfectly reveals the aggregate inventory after each round of trade. We find that the efficient number of exchanges is invariant to trading frequency, and is the same as that of the static model.

The remainder of the paper is organized as follows. Section I provides additional background on exchange market fragmentation and related research. Section II gives the setup of the most basic version of our model. Section III characterizes properties of the equilibrium. Section IV presents the implications of fragmentation on price impact, allocative efficiency, and price informativeness. Section V studies a formulation of the model in which traders observe the aggregate asset endowment before order submission. Section VI solves for the efficient number of exchanges in a dynamic formulation of the model with cross-period cross-exchange inference. Section VII summarizes the results of various model extensions. Section VIII offers some concluding remarks and discusses some potentially important effects that are not captured by our model. The online Appendix contains proofs and model extensions.

I. Background

We focus in this paper on "visible fragmentation," that is, fragmentation across different lit exchanges (meaning trade venues at which market-clearing prices are set), rather than fragmentation between lit exchanges and size-discovery

[2] A precise degree of market fragmentation that we characterize can even achieve a perfectly efficient allocation.

venues, which cross buy and sell orders at prices that are set on lit exchanges (Körber, Linton, and Vogt 2013; Zhu 2014; Degryse, de Jong, and van Kervel 2015; Duffie and Zhu 2017; Antill and Duffie forthcoming).

In Europe and the United States, exchange trading is highly fragmented. Budish, Lee, and Shim (2019) document that in the United States, as of early 2019, annual trade of about one trillion shares is split across 13 US exchanges, and that cross-exchange shares of total exchange-traded volume are stable over time, with 5 exchanges each handling over 10 percent of total exchange volume. Essentially all equities trade on every exchange, with significant volumes of each equity executed on multiple exchanges.[3] Broadly speaking, similar patterns apply to European financial markets (Gresse 2012; Degryse, de Jong, and van Kervel 2015; Foucault and Menkveld 2008). This high degree of trade fragmentation is in part a consequence of regulations such as Regulation NMS in the United States and MiFid II in Europe, which encourage exchange entry and competition.

There has been a longstanding debate (Stoll 2001) over whether fragmenting trade across exchanges harms market efficiency, in various respects. Empirical findings have been mixed (O'Hara and Ye 2011, Gomber et al. 2017). Some researchers find that fragmentation has generally been beneficial. For example, O'Hara and Ye (2011), using data from US trade reporting facilities, find that execution speeds are faster, transaction costs are lower, and prices are more efficient when the market is more fragmented. Degryse, de Jong, and van Kervel (2015) analyze a sample of Dutch stocks and measure the degree of visible fragmentation. They find that liquidity, when aggregated over all lit trading venues, improves with fragmentation. Foucault and Menkveld (2008) analyze Dutch stocks and arrive at a similar conclusion. Boehmer and Boehmer (2003) find evidence of improved liquidity when the NYSE began trading ETFs that are also listed on the American Stock Exchange. Gresse (2017); De Fontnouvelle, Fishe, and Harris (2003); Aitken, Chen, and Foley (2017); Hengelbrock and Theissen (2009); Félez-Viñas (2017); and Spankowski, Wagener, and Burghof (2012) generally find that visible fragmentation reduces bid-ask spreads.

Other research, however, suggests less beneficial effects of fragmentation. For example, Bennett and Wei (2006) find that when equity trading migrated from NASDAQ to the NYSE, where trade is more consolidated, there was a decrease in execution costs and an improvement in price efficiency. Chung and Chuwonganant (2012) show that price impact increased following the introduction of Regulation NMS.[4] Gentile and Fioravanti (2011, p. 5) find that MiFID-induced fragmentation "does not have negative effects on liquidity, but it reduces price information efficiency." Moreover, in some cases it leads primary stock exchanges to lose their leadership in the price discovery process." For small-firm equities, Gresse (2012, 2017) and Degryse, de Jong, and van Kervel (2015) find that market depth declines with sufficient fragmentation, consistent with our theoretical results. Bernales et al. (2018) find that the 2009 consolidation of Euronext's two distinct order books for the same equities was followed by a reduction in bid-offer spreads. Haslag

[3] Pagnotta and Philippon (2018) and Budish, Lee, and Shim (2019) display the striking facts graphically.

[4] In our model, as we have noted, fragmentation indeed increases price impact, yet also increases allocative efficiency and overall price informativeness.

and Ringgenberg (2020) find causal evidence that although fragmentation reduces bid-offer spreads for the equities of large firms, the opposite applies to small firms.

While the empirical evidence regarding the implications of fragmentation is mixed, most of the theoretical literature has shown that visible fragmentation is harmful. For example, Mendelson (1987) shows that fragmentation may isolate individuals for whom there are mutually beneficial trades, because they are located at different venues. Chowdhry and Nanda (1991) show that adverse selection caused by asymmetric information worsens as markets fragment. Baldauf and Mollner (2021) find that welfare is harmed by the ability of fast traders to snipe across fragmented markets. Pagano (1989) shows that fragmented markets are less stable, in that traders tend to participate at market venues at which liquidity is greatest. However, regulations promoting exchange competition may foster fragmentation.

There are few theory papers demonstrating that fragmentation may be beneficial. Of these, Kawakami (2017) shows how splitting investors across multiple exchanges can improve allocative efficiency because the resulting reduction in correlation in asset price and asset payoff can in some cases sufficiently improve the hedging effectiveness of trading. Malamud and Rostek (2017), perhaps the closest paper to ours, considers a multi-exchange demand function submission game in which each exchange operates a double auction as in our model. They find that in certain settings, when agents' risk preferences are sufficiently heterogeneous, fragmented markets can produce outcomes that are welfare superior to centralized markets.

A key difference is that Malamud and Rostek (2017) allow traders to condition their demand schedules at a given exchange on prices at other exchanges on which they participate (fully contingent demand). Because of this assumption, fragmentation may improve allocative efficiency in their model only if there is limited participation by traders across exchanges. In our model, strategic traders are allowed to submit orders to all exchanges. In their Example 2 and Proposition 4, Malamud and Rostek (2017) show a benefit of fragmentation for some asset endowments based on (i) partitioning investors across the economy's two exchanges and (ii) differences in risk aversion. Though they later show that welfare improvements may not require such an extreme form of limited participation, cross-exchange price inference, by which traders use the price in one exchange for inference regarding prices on other exchanges, does not play a role in their work as it does in ours, because they allow for fully contingent demand. For the special case of no aggregate endowment risk, Proposition 5 of Malamud and Rostek (2017) shows that fragmentation via limited participation always *lowers* welfare if there is a single asset (though this not necessarily the case when there is nonzero aggregate endowment risk or there are multiple assets and sufficient risk-aversion heterogeneity). In our model, the allocative efficiency gains from fragmentation arise for different reasons. Namely, because demand in an exchange can only be made contingent on the price within the exchange, there is no-cross exchange price impact. Thus, when splitting trade over more exchanges, the purchase of an additional unit on any given exchange affects the price on a smaller fraction of the total quantity traded. With imperfectly correlated prices, the weakening of cross-exchange price inference effects (as discussed in our introduction) ensures that price impact does not rise quickly enough with fragmentation to offset this beneficial effect of order splitting.

In practice, when mitigating price impact, strategic traders use strategies that split their "parent orders" both across exchanges and also across time.[5] In equilibrium, however, the effects of allowing more frequent trade and having more exchanges are not necessarily substitutes for improving allocative efficiency. For example, in the models of Vayanos (1999) and Duffie and Zhu (2017), increasing the frequency of opportunities to split trades worsens allocative efficiency, because equilibrium trade aggressiveness declines faster than trading frequency grows.[6] In Section VI, we allow both cross-exchange and cross-time order splitting. We find that the number E^* of exchanges that achieves an efficient reallocation of the asset at each trading date is invariant to the frequency of trade. Thus, when there are E^* exchanges (the only case that we are able to solve), more frequent trade improves allocative efficiency.

The majority of theoretical papers assume that traders are restricted to trade on a strict subset of all trading venues.[7] It seems natural to assume that traders who are strategic about their price impacts are also aware of the option to trade on multiple exchanges simultaneously. The costs of order splitting are economically small (Budish, Lee, and Shim 2019). "Smart order routing technology" makes order splitting convenient and practical (Gomber et al. 2016). In our model, strategic traders frictionlessly trade on all exchanges. There is evidence (Malinova and Park 2019, Menkveld 2008, Chakravarty et al. 2012, Gomber et al. 2016) that some investors strategically split their orders across multiple exchanges, and also between exchanges and size-discovery venues such as dark pools.

Methodologically, our model contributes to the literature on demand-function submission games, including work by Wilson (1979), Klemperer and Meyer (1989), and Malamud and Rostek (2017). Within this literature, our paper, like Wittwer (2021) and Rostek and Yoon (2020), addresses markets with multiple exchanges. While Wittwer (2021) and Rostek and Yoon (2020) focus on the welfare implications of connecting exchanges through the ability to submit orders contingent on cross-exchange prices, we consider only the common case in practice of "disconnected markets." As opposed to Wittwer (2021) and Rostek and Yoon (2020), we focus on the implications for allocative efficiency and price informativeness of increasing the number of exchanges (fragmentation), and we include a dynamic analysis that captures the implications of cross-time cross-exchange price impact, showing that enough fragmentation can achieve allocative efficiency.

Since the work of Hamilton (1979), the literature has explored the key tension between the benefit of fragmentation associated with increased competition between exchanges and between specialists, which drives down bid-offer spreads and trading fees, as suggested by the theory of Rust and Hall (2003), versus the cost of fragmentation associated with decreased market depth.[8] Although fragmentation does indeed reduce market depth in our model, consistent with earlier work, we

[5] On this point, we interviewed experts in order execution strategies at two large asset managers. Van Kervel, Kwan, and Westerholm (2020) provide evidence on order splitting behavior, by which traders with large parent orders learn over time about the presence of large parent orders of other traders.

[6] In a somewhat different setting, Du and Zhu (2017) show there can be a welfare-optimal trade frequency.

[7] For instance, Mendelson (1987), Pagano (1989), Kawakami (2017), Malamud and Rostek (2017), and many others make this assumption.

[8] For a recent empirical contribution exploring this trade-off, see Haslag and Ringgenberg (2020).

believe that we are the first to point out the benefit of fragmentation associated with increased aggregate order aggressiveness, arising from the ability of strategic traders to shield orders on a given exchange from price impacts incurred on other exchanges.

II. Baseline Model

This section presents the setup of our baseline model. All primitive random variables are defined on a complete probability space $(\Omega, \mathcal{F}, \Pr)$. There is a single asset with a payoff, denoted π, that is a finite-variance random variable with mean μ_π.

We model a market whose agents, called "traders," are of two types: "liquidity" and "strategic." For notational simplicity, we let N denote both the finite set of strategic traders and its cardinality, which is assumed to be at least 3. The only primitive information available to strategic trader i is the trader's own endowment of the asset, $X_i \sim N(0, \sigma_X^2)$. We assume that endowments are independently and identically distributed (i.i.d.) across traders.

Trade of the asset takes place in a single period on each of a finite number of identical exchanges. For notational simplicity, we let E denote both the set and number of exchanges. Each exchange runs a double auction. Strategic trader i submits a measurable demand schedule $f_{ie} : \mathbb{R}^2 \to \mathbb{R}$ to exchange e specifying the quantity $f_{ie}(X_i, p)$ of the asset demanded by trader i at any given price $p \in \mathbb{R}$ on exchange e. We emphasize that the demand schedule submitted to a given exchange cannot depend on prices (or any other information) emanating from the other exchanges. A demand schedule can be viewed as a package of limit orders, each of which is an offer to purchase or sell a given amount of the asset at a given price.[9] Liquidity traders collectively submit an exogenously given quantity $Q_e \sim N(0, \sigma_Q^2/E)$ of market orders to exchange e.

We assume that the supply of market orders is i.i.d. across exchanges and that $\{X_i \mid i \in N\}$, $\{Q_e \mid e \in E\}$, and π are independent. We relax these distributional assumptions in Section V and in extensions considered in online Appendix Section H. An interpretation of these assumptions on liquidity trade is that a large set of liquidity traders, not depending on the number of exchanges in operation, are spread evenly across exchanges and trade independently of one another.

The independence of liquidity trade across exchanges ensures that equilibrium prices are not perfectly correlated across exchanges, which is crucial for our results. If prices are perfectly correlated across exchanges, there would be no beneficial effects of fragmentation. From a practical viewpoint, the assumption that liquidity trade is not perfectly correlated across exchanges can be motivated by assuming that liquidity traders are local to an exchange, in the sense that they do not have the sophistication or trading accounts necessary to split their orders across all exchanges. Online Appendix Section H shows that it is enough for our main results to assume imperfect correlation of liquidity demands across exchanges, a weakening

[9] In this sense, $f(X_i, p)$, if positive, is the aggregate quantity of the limit orders to buy at a price of p or higher, and if negative is the aggregate quantity of the limit orders to sell at price of p or lower. The space of linear combinations of limit orders is dense, in the sense of Brown and Ross (1991), in the space of technically regular demand functions.

of independence. In a multiperiod setting, the key requirement that prices are not perfectly correlated across exchanges might be motivated by the notion that liquidity trades are not perfectly synchronized across exchanges.

Given a collection $f = \{f_{ie} \mid i \in N, e \in E\}$ of demand schedules, the price on exchange e, if it exists and is unique, is the solution[10] p_e^f to the market-clearing condition

$$\sum_{i \in N} f_{ie}\left(X_i, p_e^f\right) = Q_e. \tag{1}$$

If there does not exist a unique market clearing price, we assume that no trades are executed. We restrict attention to equilibria consisting of demand schedules with the property that p_e^f is uniquely determined.[11] Based on (1), trader i is able to determine the impact of his own demand on the market-clearing price given the conjectured demand schedules of the other traders.

The preferences of the strategic traders are quasilinear in cash compensation, with a quadratic holding cost. Specifically, given a collection $f = \{f_{ie} \mid i \in N, e \in E\}$ of demand schedules, the payoff of trader i is

$$U_i(f) = \left(X_i + \sum_e f_{ie}\left(X_i, p_e^f\right)\right)\pi - b\left(X_i + \sum_e f_{ie}\left(X_i, p_e^f\right)\right)^2 - \sum_e p_e^f f_{ie}\left(X_i, p_e^f\right),$$

for some $b > 0$. The quadratic term represents a cost for bearing the risk or other costs associated with holding a post-trade position in the asset. Preferences of this form are popular in the market microstructure literature (Vives 2011, Rostek and Weretka 2012, Du and Zhu 2017, Sannikov and Skrzypacz 2016). Sannikov and Skrzypacz (2016) provide a microfoundation.

An equilibrium is defined as a collection $f = \{f_{ie} \mid i \in N, e \in E\}$ of demand schedules with the property that, for each strategic trader i, the demand schedules $f_i = \{f_{ie} \mid e \in E\}$ solve

$$\sup_{\hat{f}} \mathbb{E}\left[U_i\left(\hat{f}, f_{-i}\right)\right],$$

where as usual f_{-i} denotes the collection $\{f_j \mid j \neq i\}$ of other traders' demand schedules. This is a typical demand-function submission game in the sense of Wilson (1979) and Klemperer and Meyer (1989), extended to allow for multiple exchanges. Multi-exchange demand function submission games were earlier analyzed by Malamud and Rostek (2017) and Wittwer (2021).

We conclude this section with an interpretation of the distinction between strategic and liquidity traders. A strategic trader may be viewed as an agent who is sophisticated, internalizes price impact, is able to easily split orders across multiple trading venues, has a relatively low aversion to owning assets, and has a relatively large initial endowment of the asset. A liquidity trader, on the other hand, may be viewed as an agent who is not sophisticated about price impacts, has high aversion to holding assets (thus exercising no discretion in the liquidation of the assets), and

[10] That is, p_e^f is a random variable such that for each state $\omega \in \Omega$, $\sum_{i \in N} f_{ie}\left(p_e^f(\omega), X_i(\omega)\right) = Q_e(\omega)$.

[11] For this, it suffices that, for each $x \in \mathbb{R}^N$, the aggregate demand function $p \mapsto \sum_i f_{ie}(p, x_i)$, which is monotone, is strictly monotone, continuous, and unbounded below and above.

has a small initial asset holding, and who therefore submits market orders with no price sensitivity. Liquidity traders are a typical modeling device for settings such as ours in which one wishes to avoid perfect inference of fundamental information from price observations. In our case, the fundamental information to be inferred does not concern asset payoffs but rather the aggregate endowment of strategic traders. Traders have payoff-relevant private information about their own endowments but no private information about asset payoffs. We will show that our main results are not driven by the effect of "donations" from liquidity traders to strategic traders.

III. A Symmetric Affine Equilibrium

We will demonstrate the existence and uniqueness of a symmetric affine equilibrium defined by demand schedules of the form

$$f_{ie}(p, X_i) \;=\; \Delta_E - \alpha_E X_i - \zeta_E p, \tag{2}$$

for constants Δ_E, α_E, and ζ_E that do not depend on the trader or particular exchange, but do depend on the number E of exchanges.

Using (1) it can be shown that the slope of the inverse residual supply curve facing each agent on each exchange is

$$\Lambda_E \;\equiv\; \frac{1}{(N-1)\,\zeta_E}. \tag{3}$$

We refer to Λ_E as inverse market depth, or simply as "price impact." Each strategic trader is aware that by deviating from the equilibrium demand schedule and demanding an additional unit on a given exchange, the trader will increase the market-clearing price on that exchange by Λ_E. Price impact is a perceived cost to each strategic trader, but is not a social cost because the payment incurred by any trader is received by another. As emphasized by Vayanos (1999), Rostek and Weretka (2015), and Du and Zhu (2017), the strategic avoidance of price impact through the "shading" of demand schedules is socially costly because it reduces the total gains from the beneficial reallocation of the asset.

Taking price impact as given, in equilibrium trader i selects a demand schedule for exchange e such that the quantity purchased equates his marginal benefit with his marginal cost for each realization of the price p_e on the exchange, in that

$$\mu_\pi - 2b\left(X_i + f_{ie}(X_i, p_e) + \mathbb{E}\left[\sum_{k \neq e} f_{ik}(X_i, p_e) \,|\, X_i, p_e\right]\right) \;=\; p_e + \Lambda_E f_{ie}(X_i, p_e). \tag{4}$$

Condition (4) would be different if each trader i could condition his demand on the realizations of prices on all exchanges on which i participates, as in Malamud and Rostek (2017). Firstly, there would be cross-exchange price impact, so the marginal cost of purchasing more of the asset would account for the effect of price impact for all units traded, as opposed to just those units traded on exchange e. Secondly, given all exchange prices, the choice by trader i of how much to purchase on exchange e would be made with perfect foresight of the quantities that he will purchase on the other exchanges. In contrast, when trader i can condition demand

only on the price on exchange e, he is uncertain about the total quantity purchased on the other exchanges, provided that the variance σ_Q^2 of liquidity trade is not zero. In order to evaluate the marginal benefit of purchasing an additional unit, he must form a conditional expectation of this quantity. As we will soon show, this gives rise to significant implications of fragmentation that are not present in a model in which demands can be conditioned on contemporaneous cross-exchange price information.

Based on the demand schedule (2), the final asset position of strategic trader i is

$$(1 - E\alpha_E)X_i + E\alpha_E \frac{\sum_{j\in N} X_j}{N} + \frac{\sum_{e\in E} Q_e}{N}. \tag{5}$$

As the market setting changes to one in which $E\alpha_E$ is higher, traders retain less of their own endowments and absorb more of the aggregate endowment. Because of this, we refer to $E\alpha_E$ as *order aggressiveness*. Indeed, increasing $E\alpha_E$ in a multi-exchange setting is allocatively equivalent in a single-exchange setting to a perception by traders of lower price impacts.

Generically in the parameters of the model, the equilibrium allocation is inefficient. Given the nondiscretionary liquidation $\sum_{e\in E} Q_e$ by liquidity traders, the efficient allocation is one in which each strategic trader receives an equal share of the aggregate supply of the asset, which is

$$\overline{q} = \frac{1}{N}\Big(\sum_{e\in E} Q_e + \sum_{i\in N} X_i\Big).$$

Inspecting (5), this efficient sharing rule corresponds to the case of $E\alpha_E = 1$. By Jensen's inequality, $E\alpha_E = 1$ yields the efficient allocation because traders have symmetric convex holding costs. Because preferences are quasi-linear in cash compensation, this is also the welfare-maximizing allocation, in that any other allocation would be strictly Pareto dominated by this efficient sharing rule, after allowing voluntary initial side payments.

The equilibrium allocation defined by (5) becomes less efficient as $|E\alpha_E - 1|$ increases. This is so because replacing $E\alpha_E$ in (5) with a number farther from 1 results in a mean-preserving spread in the cross-sectional distribution of the asset to strategic traders, state by state. Jensen's inequality, applied cross-sectionally in each state $\omega \in \Omega$, then implies an increase in the sum across traders of quadratic holding costs.

The following theorem collects several properties of symmetric affine equilibria. Of primary interest is the property that in the presence of non-trivial liquidity trade, the allocation becomes more efficient as market fragmentation E increases, up to the point at which $E\alpha_E = 1$, and then becomes increasingly less efficient. We will explore this issue in more depth in Section IV. Our proof of the theorem, found in online Appendix Section B, uses equations (2), (3), and (4) to derive a candidate set of equilibrium demand coefficients $(\Delta_E, \alpha_E, \zeta_E)$ and then applies the calculus of variations to verify that these candidate coefficients do in fact uniquely correspond to an equilibrium.

THEOREM 1: *For each positive integer number E of exchanges, there exists a unique symmetric affine equilibrium. The associated demand-function coefficients*

$(\Delta_E, \alpha_E, \zeta_E)$ form the unique solution to equations (B6), (B7), and (B8) in online Appendix Section B. Moreover:

(*i*) *For any E, the market-clearing price on exchange e is*

$$p_e^* = \frac{N-1}{N}\Lambda_E\Big[N\Delta_E - Q_e - \alpha_E \sum_{i\in N} X_i\Big]. \tag{6}$$

(*ii*) *For any E, the price impact is*

$$\Lambda_E = \frac{2b\big(1+\gamma_E(E-1)\big)}{N-2}, \tag{7}$$

where

$$\gamma_E = \frac{E\alpha_E^2\sigma_X^2(N-1)}{E\alpha_E^2\sigma_X^2(N-1)+\sigma_Q^2}. \tag{8}$$

(*iii*) *If $E > 1$, γ_E is the correlation between the prices on any two distinct exchanges from the perspective of any strategic trader i, conditional on X_i.*

(*iv*) *For any E, the final asset position of strategic trader i is given by* (5).

(*v*) *If there is no liquidity trading, in that $\sigma_Q^2 = 0$, then the equilibrium allocation does not depend on the number E of exchanges.*

(*vi*) *If $E = 1$ or $\sigma_Q^2 = 0$, then the final asset position of strategic trader i is*

$$\frac{\Lambda_1}{\Lambda_1+2b}X_i + \frac{2b}{\Lambda_1+2b}\frac{1}{N}\sum_{j\in N}X_j + \frac{\sum_{e\in E}Q_e}{N},$$

where $\Lambda_1 = 2b/(N-2)$.

(*vii*) *Suppose there is liquidity trading, in that $\sigma_Q^2 > 0$. Then the order aggressiveness $E\alpha_E$ is strictly monotone increasing in E and converges to $N/(N-1)$. In particular, a market with only one exchange is strictly dominated, from the viewpoint of allocative efficiency, by a market with any larger number of exchanges.*

Part (vi) of Theorem 1 implies that with a single exchange, the fraction of the endowment retained by a trader is increasing in the price impact Λ_1. In a centralized market, price impact avoidance is the only source of allocative inefficiency.[12] As we have described and will later elaborate, the effect of price impact avoidance on allocative efficiency can be mitigated by increasing the degree of market fragmentation. In the next section, we analyze the forces behind this and other effects of market fragmentation. But, as stated in part (vii) of Theorem 1, any degree of fragmentation is socially preferred to concentrating all trade on a single exchange.

[12] Otherwise, the proof of the First Welfare Theorem applies.

IV. The Effects of Fragmentation

We present several predictions of our model, beginning first with the effects of fragmentation on price impact.

A. *Price Impact*

Part (ii) of Theorem 1 provides the equilibrium relationship between price impact and the correlation between exchange prices. This relationship reflects the effect on trade demand of cross-exchange inference from prices. As seen from (4), the quantity $f_{ie}(X_i, p_e)$ purchased by trader i on exchange e at the price p_e depends on the probability distribution of the quantities that trader i will execute on the other exchanges, conditional on X_i and p_e.

To illustrate, suppose for example that in state $\omega \in \Omega$ trader i is a buyer of the asset at the equilibrium price on exchange e. If the observed price outcome $p_e(\omega)$ was lowered, trader i would assign a higher conditional likelihood to lower prices on the other exchanges because strategic traders' demands are positively correlated on any two exchanges (which implies a positive cross-exchange price correlation, γ_E). But trader i submits demands to the other exchanges before observing p_e. Thus, the lower is $p_e(\omega)$, the higher is the conditional expected quantity executed by trader i on the other exchanges. As $p_e(\omega)$ declines, the marginal utility of trader i per unit purchased on exchange e is also reduced. Due to cross-exchange inference, the quantity purchased by trader i on exchange e in response to a decrease in price $p_e(\omega)$ is smaller than would be the case if there was no cross-exchange correlation. Analogous reasoning can be applied to show that, due to cross-exchange inference, the quantity purchased by trader i optimally on exchange e in response to an increase in the price $p_e(\omega)$ is smaller than it would be if there was no cross-exchange correlation. Overall, cross-exchange price inference reduces the steepness (absolute slope) of the demand schedule of trader i on each exchange. The result, by (3), is that price impact rises. Because this channel does not exist with a single centralized exchange, price impact is always higher in a fragmented market than in a centralized market.

We now discuss comparative-static effects for price impact associated with a change in the variance σ_Q^2 of liquidity trade and the number E of exchanges. As σ_Q^2 increases, prices at different exchanges become less correlated, so price impact declines, eventually converging to that of a single exchange market as σ_Q^2 tends to infinity. Thus, price impact is lower in markets with noisier liquidity trader supply because the cross-exchange inference channel is weaker. The following proposition characterizes the dependence of price impact on the number of exchanges.

PROPOSITION 1: *The price-impact coefficient Λ_E is strictly monotone increasing in the number E of exchanges. If the variance σ_Q^2 of liquidity trade is zero, then $\lim_{E\to\infty}\Lambda_E = \infty$. If $\sigma_Q^2 > 0$, then*

$$\lim_{E\to\infty}\Lambda_E = \frac{2b}{N-2}\left(1+\frac{N^2\sigma_X^2}{(N-1)\sigma_Q^2}\right),$$

and for $E > 1$, cross-exchange price correlation γ_E declines strictly monotonically as E increases and converges to zero as $E \to \infty$.

Proposition 1 states that market fragmentation increases price impact and (with nontrivial liquidity trade) reduces cross-exchange price correlation. Without liquidity trade ($\sigma_Q^2 = 0$), price impact diverges as the number of exchanges diverges, because γ_E is equal to 1. But with liquidity trade ($\sigma_Q^2 > 0$), price impact converges to a finite value. Because price impact depends on $\gamma_E(E-1)$, this follows from the fact that γ_E declines at a rate proportional to $1/E$. The intuition is that as the number of exchanges increases, the expected quantity traded on a given exchange decays at rate $1/E$, which in turn causes the variability in prices due to strategic traders' orders to decay at a rate proportional to $1/E^2$. Since the variability in prices due to exchange-specific liquidity trade is σ_Q^2/E, this implies that γ_E must decline at the rate $1/E$, so that price impact converges.

Figure 1 illustrates the relationship between price impact and the number of exchanges, for different cases of the number N of strategic traders. As illustrated, price impact converges faster when there are more strategic traders. For instance, consider the case of $b = 1/2, N = 5$, and $E = 100$. Without liquidity trade, price impact is roughly $\Lambda_E = 33$. However, with $\sigma_Q^2 > 0$ and strategic traders whose endowments are 10 times more uncertain (in terms of variance) than aggregate liquidity trader supply (in that $\sigma_X^2/\sigma_Q^2 = 10$), price impact drops to approximately 10. As σ_X^2/σ_Q^2 falls below 10, γ_E is reduced and, because of this, price impact is further reduced.

B. *Allocative Efficiency*

We have just shown that price impact is higher in more fragmented markets. However, Theorem 1 tells us that, provided there is no liquidity trade ($\sigma_Q^2 = 0$), even though price impact diverges as E tends to infinity, total trade aggressiveness is unaffected and the equilibrium allocation remains constant. Moreover, when $\sigma_Q^2 > 0$, even though price impact increases with fragmentation, total trade aggressiveness actually *increases*. One might have expected that the rise in price impact would lead to a reduction in trade aggressiveness and thus lower allocative efficiency, but this is not the case. We turn now to a resolution of this superficial paradox.

As fragmentation rises, price impact rises but traders are better able to evade the overall cost of price impact by splitting their orders across exchanges. This is so because traders bear the cost of price impact on a given exchange only to the extent of the trades executed on that exchange. By order splitting, a trader can shield an order on a given exchange from the price impact of units executed on the other exchanges. When there are more exchanges, the purchase of an additional unit on a given exchange affects a smaller fraction of the total quantity traded. When there is no liquidity trade ($\sigma_Q^2 = 0$) this effect exactly offsets the rise in price impact, leaving the overall aggressiveness of a trader's demand invariant to the number of exchanges. When $\sigma_Q^2 > 0$, price impact does not rise quickly enough to offset the effect of increased aggressiveness through order splitting.

As a source of intuition, suppose that trader i purchases q/E additional units of the asset on each exchange, for a total of q units. Because there is no cross-exchange

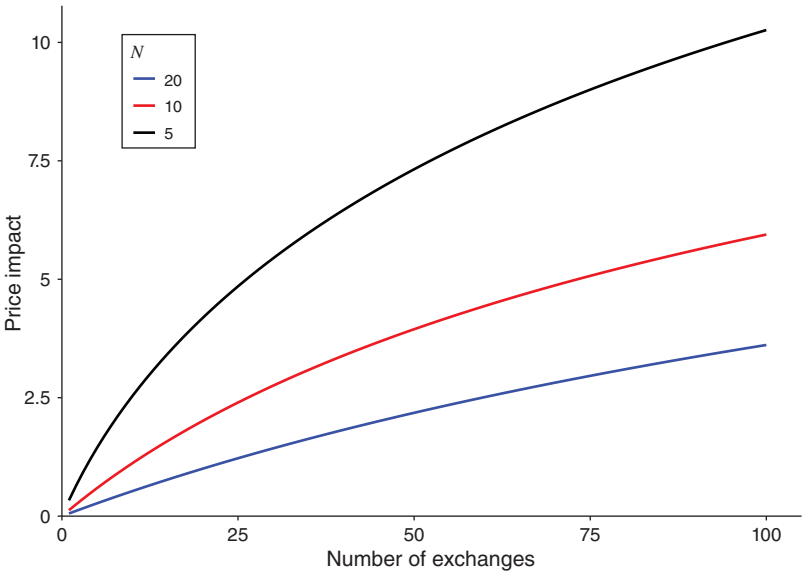

FIGURE 1. PRICE IMPACT

Notes: Variation of price impact Λ_E with the number E of exchanges, for various cases of N, the number of strategic traders. In all cases, the variance-aversion coefficient is $b = 1/2$ and the ratio σ_X^2/σ_Q^2 of the variance of strategic-trader asset endowment to total liquidity trade quantity is 10.

price impact, this increases the price on each exchange by $\Lambda_E q/E$. Thus, the overall price-impact cost to trader i is $\Lambda_E q^2/E$. If the price impact coefficient Λ_E were to remain constant, an increase in the number of exchanges would reduce the overall cost of price impact, because of order-splitting. Price impact in fact rises with E, but overall order aggressiveness depends on Λ_E/E. By part (ii) of Theorem 1,

$$\frac{\Lambda_E}{E} = \frac{2b}{N-2}\left(\gamma_E + \frac{1-\gamma_E}{E}\right).$$

If the equilibrium cross-exchange price correlation, γ_E, is imperfect and non-increasing in E, then $\Lambda_E q/E$ is decreasing[13] in E. In fact, Proposition 1 states that if $\sigma_Q^2 > 0$, then γ_E is strictly decreasing in E and ultimately converges to zero. Intuitively, when prices are imperfectly correlated, the effects of cross-exchange inference discussed in Section IVA are weakened by market fragmentation, so that price impact rises less quickly than the number of exchanges. Thus, the dominant effect on overall aggressiveness is order-splitting, and overall price impact costs decline with E. As a result, agents trade more aggressively in more fragmented markets.

At low levels of fragmentation, increases in trade aggressiveness are beneficial for allocative efficiency. But when markets become sufficiently fragmented, additional

[13] The assumption that liquidity trade is independent across exchanges ensures $\gamma_E < 1$. If liquidity trades were instead perfectly correlated across exchanges, then γ_E would be 1 and fragmentation would have no effects on allocative efficiency.

increases in aggressiveness are inefficient, in that $E\alpha_E$ increases past the point of efficiency, at which $E\alpha_E = 1$, rising with ever larger E to the limit $N/N-1$. We emphasize, however, that trade never becomes so aggressive that fragmentation leads to a loss of allocative efficiency relative to that of a market with a single exchange.

By equation (5), the number of exchanges that maximizes allocative efficiency is that for which $E\alpha_E$ is closest to 1.

PROPOSITION 2: *Suppose* $\sigma_Q^2 > 0$. *Let*

$$E^* = 2 + \frac{2}{N-2} + \frac{N-1}{N-2}\frac{N\sigma_X^2}{\sigma_Q^2}. \tag{9}$$

If E^* *is an integer, the unique symmetric affine equilibrium for a market with* E^* *exchanges achieves an efficient allocation of the asset, by allocating an equal amount* $\bar{q}$ *of the asset to each strategic trader. In general, the number of exchanges that maximizes allocative efficiency is* $\lfloor E^* \rfloor$ *or* $\lceil E^* \rceil$.

By Proposition 2, the optimal number of exchanges is finite, is at least 2, and depends crucially on the ratio of the variance of the endowment of strategic traders to the variance of the total amount of liquidity trade, σ_X^2/σ_Q^2. This ratio determines the cross-exchange price correlation γ_E, as seen in equation (8), which in turn determines price impact. As σ_X^2/σ_Q^2 rises, price impact is higher, so more fragmentation necessary to offset the adverse effect of price impact with the beneficial effect of increasing the number of exchanges over which strategic traders can split their orders.

It is perhaps surprising that the socially optimal number of exchanges is finite. For any finite E, price impact costs are positive and only disappear in the limit, in that $\lim_{E\to\infty} \Lambda_E/E = 0$. It turns out, however, that fragmentation introduces an additional inefficiency beyond price impact. For intuition, consider again equation (4), which determines the demand of trader i at a given price, p_e on exchange e. To evaluate the marginal benefit of purchasing additional units, trader i must form a conditional expectation of the quantities that he will purchase on the other exchanges. To form this expectation, trader i must in turn form a conditional expectation of the prices on the other exchanges. If prices at distinct exchanges are perfectly correlated, then $\mathbb{E}[p_k|X_i,p_e] = p_e$ for each exchange k. However, if exchange prices are imperfectly correlated, then, using (6), it can be shown that

$$\mathbb{E}[p_k \,|\, X_i,p_e] = -(1-\gamma_E)\frac{N-1}{N}\Lambda_E(\alpha_E X_i - N\Delta_E) + \gamma_E p_e, \tag{10}$$

for an arbitrary exchange $k \neq e$. That is, trader i must use his own endowment X_i to forecast prices on the other exchanges. The effect of using X_i for cross-exchange inference is an increase in α_E, the sensitivity of demand with respect to X_i.

The intuition behind this is analogous to that of Section IVA, where we explained why cross-exchange inference leads to a rise in price impact. Holding p_e fixed, if X_i is lowered, trader i would expect prices on the other exchanges to rise, as seen by (10), and thus would expect to purchase less on the other exchanges. The marginal benefit of purchasing an additional unit on exchange e would then rise.

As a result, the sensitivity of the quantity purchased by trader i to reductions in X_i is increased by cross-exchange inference. That is, total order aggressiveness α_E increases because of the role of X_i in cross-exchange inference. This force is not present in a market with a single exchange. With a single exchange, the allocation would be efficient if traders ignored their price impact. In contrast, if $\sigma_Q^2 > 0$, then as E tends to infinity, even though price impact disappears, trade eventually becomes overly aggressive from a welfare perspective.

Figure 2 illustrates the results of this section. As shown, $E\alpha_E$ is strictly increasing in fragmentation and can exceed the socially efficient level. The socially efficient number of exchanges increases with σ_X^2/σ_Q^2.

C. *Price Informativeness*

Our finding that trade aggressiveness increases with market fragmentation has natural implications for price informativeness. By price informativeness, we mean the degree to which prices reveal information about the average endowment $\bar{X} = \sum_{i \in N} X_i/N$ of strategic traders. This notion is especially relevant when viewing our model as though a snapshot of a dynamic market in which liquidity trade is serially uncorrelated and the aggregate strategic endowment is a persistent Markov process. In such a setting, the aggregate endowment of strategic traders is a sufficient statistic for inference regarding future prices and future aggregate endowments.

Because of the joint normality of prices and endowments in our model, the conditional variance of $\bar{X}$ given exchange prices is an unambiguous metric for price informativeness. Our results are summarized in Proposition 3.

PROPOSITION 3: *Suppose that the variance σ_Q^2 of liquidity trade is not zero. Then:*

(*i*) *For any exchange e, $\mathrm{var}\left(\bar{X}|p_e^*\right)$ is strictly monotone increasing in the number E of exchanges and converges to $\mathrm{var}\left(\bar{X}\right)$ as E goes to ∞.*

(*ii*) $\mathrm{var}\left(\bar{X}|\{p_e^* : e \in E\}\right)$ *is strictly monotone decreasing in E.*

That is, as market fragmentation rises, the informativeness of the price on any individual exchange worsens, but overall price informativeness, taking into consideration information from all exchange prices, improves.

V. The Case of Observable Aggregate Endowment

This section presents a simplified version of the model in which the aggregate endowment of strategic traders is publicly observable. This allows a demonstration of the welfare benefits of fragmentation in a setting that requires neither liquidity traders nor Gaussian asset endowments. As before, the equilibrium price on a given exchange is linear with respect to the aggregate endowment and exchange-specific liquidity trade. Thus, conditional on the aggregate asset endowment which we now assume is publicly observable, prices on any two exchanges are uncorrelated, so traders do not make cross-exchange inferences from prices. This shuts down the

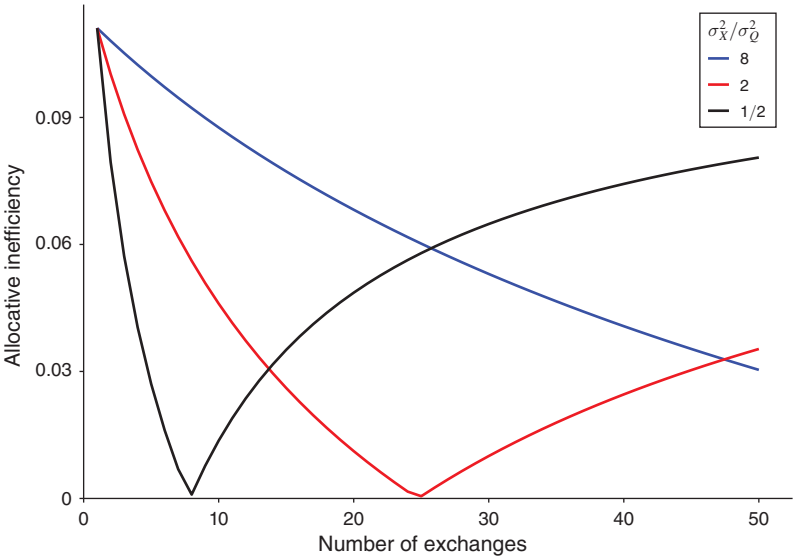

FIGURE 2. ALLOCATIVE INEFFICIENCY

Notes: We plot equilibrium allocative inefficiency as measured by $|1 - E\alpha_E|$ against the number of exchanges for different values of the ratio σ_X^2/σ_Q^2 of the variance of the endowment of a strategic trader to the variance of the total amount of liquidity trade. In all cases, the number N of strategic traders is 10. Allocative inefficiency, $|1 - E\alpha_E|$, does not depend on the variance-aversion coefficient b.

cross-exchange inference channel, allowing an isolated analysis of the welfare benefits of order splitting.

We retain the model setup of Section II with the exceptions that, for any exchange e and any trader i, (i) Q_e and X_i are of finite variance but not necessarily normally distributed (though Q_e still has mean zero) and (ii) trader i observes[14] the private endowment X_i and the average endowment $\bar{X}$. The following theorem characterizes the equilibrium of this model.

THEOREM 2: *For each number E of exchanges, there exists a symmetric affine equilibrium. If, in addition, for each e, Q_e has full support on $\mathbb{R}$, then the equilibrium is unique in the class of symmetric affine equilibria, and:*

(*i*) *The price-impact coefficient $\Lambda_E = 2b/(N-2)$ does not depend on the number E of exchanges.*

[14] That is, the demand submitted by trader i on exchange e is a measurable function $f_{ie}: \mathbb{R}^3 \to \mathbb{R}$ that, at any price p, determines the demand $f_{ie}(X_i, \bar{X}, p)$.

(*ii*) *For each fixed number E of exchanges:*

(*a*) *The price on exchange e is*

$$p_e^* \;=\; -2b\Big(\bar{X} + \frac{Q_e}{N-2}\frac{N-1}{N}\Big) + \mu_\pi.$$

(*b*) *The final asset position of trader i is*

$$\frac{\Lambda_E}{\Lambda_E + 2bE}X_i + \frac{2bE}{\Lambda_E + 2bE}\bar{X} + \frac{\sum_{e\in E} Q_e}{N}.$$

(*iii*) *Allocative efficiency is increasing in the number E of exchanges. As E diverges, the allocation converges to the efficient allocation,* $\bar{q}$ *to each strategic trader.*

(*iv*) *The total expected equilibrium payment* $-\mathbb{E}\big[\sum_{e\in E} p_e^* Q_e\big]$ *of liquidity traders is invariant to the number E of exchanges and is equal to*

$$\frac{2b\,\mathrm{var}\big(\sum_{e\in E} Q_e\big)}{N-2}\frac{N-1}{N}.$$

In this setting, price impact is a constant that does not depend on the level of fragmentation because there is no cross-exchange inference effect. By part (iii) of the theorem, more fragmentation is unambiguously beneficial in this setting. In the limit as E tends to infinity, the fully efficient allocation obtains. The benefits here of fragmentation arise entirely from the beneficial effects of increased order aggressiveness associated with order splitting. An equilibrium exists with the properties stated in the theorem even if there is no liquidity trade, in that $Q_e = 0$ for each e. However, liquidity trade is still needed for equilibrium uniqueness. Even in the presence of liquidity traders, the total expected payment of liquidity traders is invariant to market fragmentation. Thus, the beneficial effect of fragmentation is not related to the exploitation of liquidity traders by strategic traders.[15] In the model of Section II, the liquidity traders are only a convenient modeling device for breaking the perfect correlation in exchange prices. In an alternative high-frequency multiperiod setting, Budish, Cramton, and Shim (2015) note that the prices of similar assets on different exchanges are virtually uncorrelated, empirically. We explore a multi-period setting in the next section.

VI. A Dynamic Model

One might guess that market fragmentation would not support allocative efficiency as well in a dynamic setting as in a static setting. In our static setting, the

[15] In the setting of Section IV, our results are not driven by donations from liquidity traders, but liquidity traders do pay more in expectation as fragmentation increases. The total expected payment to strategic traders is

$$\mathbb{E}\Big[\sum_{e\in E} p_e^* Q_e\Big] \;=\; \frac{N-1}{N}\Lambda_E \sigma_Q^2, \tag{11}$$

which is strictly increasing in E since Λ_E is strictly increasing.

price impact of a trader's orders on exchange e does not increase trading costs on other exchanges, because of the simultaneity of trade. In a dynamic setting, however, when submitting a trade on exchange e at period t, a trader internalizes the resulting impact on the prices on all exchanges in period $t+1$, given the inference about aggregate inventory that is drawn by other traders from observing p_{et}. Nevertheless, in this section, we show that market fragmentation allows efficient trade even in a dynamic setting, despite the associated cross-period cross-exchange price impact and higher within-period price impact. Moreover, the efficient number of exchanges is invariant to trading frequency, and is the same as that of the static model.

A. *Setup*

Trade occurs at each of a discrete set of times separated by some duration Δ. A positive integer t denotes the tth trading date. As in the baseline static model, E exchanges operate separate double auctions for a single asset at each trade time. The asset pays π_t at date t, post-trade, where $\pi_1, \pi_2, \ldots$ are independent with common mean $\mu_\pi \Delta$. Liquidity traders supply a Gaussian quantity Q_{et} of the asset to exchange e at trade date t, independent across exchanges and dates, with common mean zero and variance $\sigma_Q^2 \Delta$. Immediately prior to trade at date t, trader i receives a Gaussian inventory shock ϵ_{it} that has mean zero and variance $\sigma_X^2 \Delta$, independent across trading dates and traders. The inventory shocks, liquidity trader supplies, and the asset payoffs are independent.

The pre-trade inventory of trader i at period t is

$$X_{it} = X_{i,t-1} + \sum_{e \in E} q_{ie,t-1} + \epsilon_{it}, \tag{12}$$

where q_{iet} is the quantity purchased by trader i on exchange e at period t. For $t = 0$, we set $X_{i,t-1} + \sum_{e \in E} q_{ie,t-1} = 0$. Note that since the equilibrium we construct has efficient trade, the variance of ϵ_{it}, $\sigma_X^2 \Delta$, is also the *conditional* variance of X_{it} from the perspective of each trader following trade at date $t-1$. It is not, however, the unconditional variance of X_{it} for $t > 0$.

During the time interval $\big[t\Delta, (t+1)\Delta\big)$, the net payoff to trader i, discounted to the beginning of the interval at the rate $r > 0$, is the total initial payoff from asset holdings, net of asset purchase costs, plus discounted inventory holding costs, given by

$$\begin{aligned} F_{it}(q_{it}) &= \pi_t\Big(X_{it} + \sum_{e \in E} q_{iet}\Big) - \sum_{e \in E} p_{et} q_{iet} - \int_0^\Delta e^{-rs} \tilde{b} \Big(X_{it} + \sum_{e \in E} q_{iet}\Big)^2 ds \\ &= \pi_t\Big(X_{it} + \sum_{e \in E} q_{iet}\Big) - \sum_{e \in E} p_{et} q_{iet} - b\Big(X_{it} + \sum_{e \in E} q_{iet}\Big)^2, \end{aligned} \tag{13}$$

where $q_{it} = (q_{i1t}, \ldots, q_{iEt})$ and

$$b = \tilde{b} \frac{1 - e^{-r\Delta}}{r}.$$

Our formulation is in the spirit of Vayanos (1999), differing mainly in that we allow multiple exchanges, introduce liquidity traders, and assume a different

inventory preference model. For tractability, a significant part of the analysis in Vayanos (1999) focuses on the case in which σ_X^2 tends to zero. Our analysis applies to arbitrary σ_X^2.

We do not solve for an equilibrium of the model for an arbitrary number E of exchanges because of the problem of infinite regress of beliefs, as described in the conclusion of Vayanos (1999). In the presence of liquidity traders, strategic traders choose their trades based on their beliefs about the aggregate market asset inventory, as well as beliefs about other traders' beliefs about aggregate inventory, beliefs about the beliefs of other traders about their own beliefs, and so on, causing the state space to explode. To our knowledge, there has been no analysis of demand-function submission games in which traders filter information from prices so as to discern strategic trading from liquidity trading. Rather than addressing equilibria for general E, we instead construct an equilibrium for a specific number E of exchanges with the property that the associated equilibrium is perfect Bayesian and implements efficient trade. This equilibrium is tractable because efficient trade dramatically simplifies the inference problem of each trader, given that the sum of exchange prices perfectly reveals the aggregate inventory after each round of trade.

B. *An Equilibrium with Efficient Trade*

In this section we briefly derive an efficient equilibrium and characterize its key properties, including the associated number E of exchanges. A proof is given in online Appendix Section E.

To start, we conjecture that there exists a number E of exchanges such that in equilibrium each trader i submits the demand schedule to exchange e given by

$$f_{iet}(X_{it}, p_{et}, B_t) = -\frac{1}{E}X_{it} - \zeta p_{et} + \rho B_t + \chi, \tag{14}$$

for some constants ζ, ρ, and χ to be determined, where B_t is defined recursively by $B_0 = 0$ and

$$B_t = NE\rho B_{t-1} + NE\chi - \zeta N \sum_{e \in E} p_{e,t-1}. \tag{15}$$

We later interpret B_t as a variable related to trader beliefs about the aggregate supply of the asset. Given the conjectured form (14) of the demand function, market clearing implies that the equilibrium price on exchange e is

$$p_{et} = \frac{N\rho B_t + N\chi - Q_{et} - \frac{1}{E}\sum_{j \in N} X_{jt}}{\zeta N}. \tag{16}$$

The post-trade aggregate inventory of strategic traders at date t is

$$W_t = \sum_{j \in N} X_{jt} + \sum_{e \in E} Q_{et}.$$

Substituting (16) into (14) and summing across $e \in E$, we verify that the final inventory of trader i at date t is efficient and equal to W_t/N. By substituting (16) into (15)

we see that, along the equilibrium path, B_t is equal to W_{t-1}. If any given trader were to deviate from the equilibrium strategy prior to date t, it is possible that $B_t \neq W_{t-1}$. Nonetheless, even if traders had deviated prior to date t, any trader who has not deviated must believe that $W_{t-1} = B_t$ with probability 1 because the Gaussian liquidity trading and inventory shocks ensure that deviations are undetectable. Thus, any given trader i must believe that any other trader j believes that $W_{t-1} = B_t$, and so on with respect to higher-order beliefs. Thus, B_t is a sufficient statistic for higher-order beliefs. This allows for a tractable equilibrium construction.

We now provide intuition for the role of the key state variable B_t in traders' demand schedules. If trader j follows the equilibrium strategy, (15) and (14) imply that

$$X_{j,t-1} + \sum_{e \in E} q_{je,t-1} = \frac{1}{N} B_t.$$

Summing across $j \neq i$,

$$\sum_{j \neq i} \left(X_{j,t-1} + \sum_{e \in E} q_{je,t-1} \right) = \frac{N-1}{N} B_t.$$

Thus for trader i, B_t is a sufficient statistic for the total post-trade inventory of other traders at date $t-1$. This in turn implies that B_t is sufficient information for trader i to conduct inference about the residual supply that he will face on each of the exchanges at time t. This explains the role of B_t in the demand schedule (14).

In a perfect Bayesian equilibrium, any given trader i, conjecturing that other traders submit demand functions according to (14), solves the stochastic control problem

$$\sup_{\{f_{iet}\}} \mathbb{E} \left[\sum_{t=0}^{\infty} e^{-r\Delta t} F_{it}(q_{it}) \,\middle|\, X_{i0} \right], \tag{17}$$

with demands that are measurable[16] with respect to the history of inventory levels $\{X_{is}\}_{s \leq t}$, trades $\{q_{ies}\}_{e \in E, s<t}$, and prices $\{p_{es}\}_{e \in E, s<t}$, and satisfy[17]

$$\lim_{t \to \infty} e^{-r\Delta t} \mathbb{E}\left[X_{it}^2\right] = 0, \tag{18}$$

ruling out "Ponzi schemes" that are based on explosive growth in asset positions. An equilibrium is characterized by optimal demands determined by the same function $f_{iet}(\cdot)$ of (14).

In solving the optimization problem (17), trader i correctly considers the impacts of his trades on current and future prices. These impacts occur directly through the formation of the clearing price on the exchange to which an order is submitted and also through the recognition by trader i that other traders draw inference from

[16] Although the objective function involves second moments of X_{it}, we allow strategies that do not have finite second moments and show that any such strategy is strictly suboptimal.

[17] This condition is implied by the square-integrability condition $\mathbb{E}\left[\sum_{t=0}^{\infty} e^{-r\Delta t} \sum_{e \in E} q_{iet}^2\right] < \infty$.

market prices about the aggregate market supply of the asset, which affects future prices at all exchanges. This impact occurs through the "beliefs" state variable B_t, through the dynamic equation (15).

In online Appendix Section E, we use Bellman's principle of optimality to calculate the required number E^* of exchanges, which is given by (9), the same as that for the static model. Solutions for the equilibrium demand coefficients ρ, ζ, and χ are reported at the end of online Appendix Section E.

From the demand schedule (14), the within-period price impact on any exchange is

$$\frac{1}{\zeta(N-1)} = \frac{2b\big(1+\Gamma(E-1)\big)}{N-2-e^{-r\Delta}\frac{2N-2}{E}\big(1+\Gamma(E-1)\big)}, \tag{19}$$

which is higher than in the associated static model. We also compute the cross-period cross-exchange price impact

$$\frac{dp_{e,t+1}}{dq_{jkt}} = -N\rho\frac{1}{(N-1)\zeta} = \frac{1}{E}\frac{2b\big(1+\Gamma(E-1)\big)}{N-2-e^{-r\Delta}\frac{2N-2}{E}\big(1+\Gamma(E-1)\big)}, \tag{20}$$

which is a fraction $1/E$ of the within-period within-exchange price impact. (Our differential notation for this price sensitivity involves a transparent abuse of notation.) The marginal impact of the quantity traded by any trader on any exchange on the sum of exchange prices in the next time period is equal to the within-period within-exchange price impact.

C. *Summary of Results*

The following theorem summarizes the results of our analysis of the dynamic model.

THEOREM 3: *If E^* of (9) is an integer, then there exists a perfect Bayesian equilibrium in symmetric affine demand schedules for the dynamic market with E^* exchanges such that:*

(*i*) *Trade is allocatively efficient along the equilibrium path.*

(*ii*) *Traders submit the demand schedule given by (14), with ρ, ζ, and χ given by (E23), (E24), and (E25), respectively.*

(*iii*) *Beliefs about the aggregate market inventory evolve according to (15).*

(*iv*) *Trades on each exchange have nonzero price impact at each exchange in the next period, given by (20).*

(*v*) *The within-period within-exchange price impact (19) is higher than that for the associated static model.*

In the equilibrium, by deviating, traders can manipulate other traders' beliefs about the aggregate market asset inventory. Following a one-shot deviation, trade returns to efficiency in the next period, and beliefs become "corrected."

Our analysis shows that in our dynamic model, as for the associated static model, a precise and nontrivial amount of market fragmentation achieves allocative efficiency. Relative to the static model, our dynamic model allows a clearer characterization of how fragmentation improves price discovery. A weakness of our analysis of the dynamic setting is that, because of the need to incorporate the infinite regress of beliefs about beliefs, we are able to characterize equilibrium only for the number of exchanges that is associated with an efficient allocation. A notable implication of Theorem 3 is that this efficient number of exchanges is invariant to the frequency of trade and is identical to that of the static model. Though the presence of multiple periods of trade increases price impact relative to the static model, it also amplifies the role of traders' asymmetric information, which, as we saw in the static model, leads to more aggressive order submission. This is so because traders rely on their privately observed inventories to conduct inference that is relevant not only to current-period trade prospects, but also to future-period trade prospects. This effect on trade aggressiveness precisely offsets the effects of the rise in price impact on each exchange.

Our analysis has side implications for the optimal trading frequency. We show in online Appendix Section E that our model is essentially equivalent to a setting in which traders' inventories are continually shocked by independent Brownian Motions. In that setting, since Theorem 3 also applies, for the efficient number E^* of exchanges, allocative efficiency increases as trade frequency increases. This contrasts with Vayanos (1999) and Du and Zhu (2017), who show that with a single exchange, allocative efficiency may decline as trade frequency increases.

VII. Discussion of Model Extensions

In this section we summarize the results of three extensions of the main model that are provided in the online appendices.

A. *Endogenous Liquidity Trade, Exchange by Exchange*

In our first model extension, found in online Appendix Section F, liquidity traders, who are local to each exchange and conduct no cross-exchange trade, choose the sizes of their trades. Liquidity traders are assumed to have the same preferences as strategic traders, with the exception of a potentially different quadratic holding cost parameter, c. They are also each endowed with a Gaussian distributed quantity of the asset prior to trade. Thus, the baseline model is equivalent to the case in which $c = \infty$, in that liquidity traders liquidate their entire endowed positions as though without discretion. Relaxing this baseline extreme assumption to the case of finite c, we find for any positive integer $\bar{E} > 1$, there exists a cutoff $\bar{c}$ such that if $c > \bar{c}$, then a market with $1 < E \leq \bar{E}$ exchanges is welfare superior to a centralized market, in that the expected sum of all agents' holding costs is lower.

B. *Private Information about Asset Payoff*

In a second extension, found in online Appendix Section G, agents have differing private information about the asset's final payoff. In this case, allocative efficiency is not necessarily improved by fragmenting a centralized market. This is so because fragmentation leads agents to trade more aggressively for two reasons: not only to mitigate holding costs, but also to exploit payoff-relevant private information. While the former motive leads fragmentation to improve allocative efficiency, as we demonstrated in Section IV, the latter effect can cause fragmentation to reduce allocative efficiency, because the efficient allocation of the asset does not depend on agents' payoff-relevant private information. Whether fragmentation is beneficial or harmful is shown to depend on the relative magnitudes of these two effects.

C. *Correlated Trade Motives*

In a third extension, found in online Appendix Section H, we relax the assumption that the asset endowments $(X_1, \ldots, X_N, Q_1, \ldots, Q_E)$ are jointly independent. We retain the assumption that these random variables are jointly Gaussian, but allow for an essentially arbitrary covariance matrix, subject to the condition that the traders' endowments $X_1, \ldots, X_N$ are symmetrically distributed and that the liquidity-trade quantities $Q_1, \ldots, Q_E$ are symmetrically distributed.

If a strategic trader's endowment X_i does not covary more negatively with aggregate liquidity trader supply $\sum_e Q_e$ than it covaries positively with the aggregate endowment $\sum_j X_j$, then there is an interior optimal level of fragmentation which, up to the integer constraint on E, achieves the efficient allocation.[18]

In this setting, however, an arbitrary level of market fragmentation need not coincide with an unambiguous improvement in allocative efficiency over a centralized market. Whether this is so depends on the covariances of endowments. With some parameters, agents may trade even more aggressively than they do in the baseline model, which we have shown has the property that trade already becomes "too aggressive" for sufficiently large E. Moreover, if a strategic trader's endowment covaries more negatively with the aggregate liquidity trader supply than it covaries positively with the aggregate endowment, fragmentation is harmful. This is so because the inefficiency associated with the inferior trading technology associated with disconnected fragmented markets dominates the beneficial effect of reducing the strategic avoidance of price impact. This follows from the fact that, ex ante, with this correlation structure, traders expect that the residual supply on each exchange is on average relatively favorable for offsetting their positions. This, however, leads to less aggressive trade than is socially efficient since agents are less willing to trade large quantities at unfavorable prices on any given exchange because they expect that prices on the other exchanges will be more favorable.

[18] Positive definiteness of the covariance matrix ensures, for each i, X_i is positively correlated with $\sum_j X_j$.

VIII. Concluding Discussion

We have presented a simple market setting in which fragmentation of trade across multiple exchanges improves allocative efficiency and price informativeness. Our main marginal contributions are (i) a newly identified channel by which cross-exchange price inference exacerbates price impact, and (ii) a demonstration of the beneficial effects of cross-exchange order-splitting on allocative efficiency and price informativeness. We find that although fragmentation reduces market depth on any given exchange, this need not be a sign of worsening overall liquidity or market inefficiency. We characterize the number of exchanges that achieves allocative efficiency, and show that this "optimal" degree of fragmentation is invariant to the frequency of trade and indeed the same as that of the static version of the model.

Our stylized model abstracts from many important practical considerations. We do not consider some of the direct frictional costs of trade and order splitting, such as trading fees and subsidies, minimum tick sizes, and bid-offer spreads, which are endogenous to market structure, particularly through the role of competition among exchange operators, specialists, and market makers (Baldauf and Mollner 2020; Chao, Yao, and Ye 2018; Colliard and Foucault 2012; Malinova and Park 2019; Foucault and Menkveld 2008; Chlistalla and Lutat 2011; Clapham et al. 2021; Hengelbrock and Theissen 2009; Parlour and Seppi 2003). For example, Foucault and Menkveld (2008) show that, with nonzero tick sizes, adding an additional limit-order market increases market depth by allowing limit-order submitters to jump the queue of posted orders on one exchange by posting orders on another exchange, due to the absence of cross-exchange time-priority rules. Foley, Jarnecic, and Liu (2020) show that liquidity providers increasingly fragment their activities amongst alternative venues, attempting to jump long queues on larger venues by increasing submissions to venues with short (or empty) queues. This reduces adverse selection costs faced on alternative venues and helps explain the increase in fragmentation for jurisdictions with trade-through prohibitions.

We also do not consider the endogenous entry of exchanges, a common theme in the literature going back to Glosten (1994), as reviewed by Pagnotta and Philippon (2018). Our model does not capture the effect of high-frequency traders who can take advantage of slight discrepancies in order execution times across different exchanges (Budish, Lee, and Shim 2019; Gresse 2012; Pagnotta and Philippon 2018). We also ignore the role of trade-through rules such as Regulation NMS, which effectively forces all US lit exchanges to recognize the best bid or offer available across all order books in the market. While Reg NMS has the effect of consolidating markets for small trades, trade-through rules do not play a significant role in price-impact costs, which are only pertinent for large trades. The inefficiencies associated with price-impact cost avoidance through order splitting are the main concern in this paper.

Because we have abstracted from these and other potentially important realistic effects, we make no normative claims or policy recommendations. The mechanisms that we identify do, however, appear to have a natural basis and to be worthy of serious consideration in policy discussions.

Our model also has implications for the welfare impact of innovation of trading technologies. For example, the beneficial welfare effects of order splitting that we

have described rely crucially on the realistic assumption that orders submitted to one exchange cannot condition on prices at other exchanges. If, instead, trading technology were to allow orders to condition on cross-exchange prices, then trades on a given exchange would have impact on prices at other exchanges, which could eliminate the beneficial effect of order-splitting in fragmented markets, an issue considered by Wittwer (2021) and Rostek and Yoon (2020).

REFERENCES

Aitken, Michael, Haoming Chen, and Sean Foley. 2017. "The Impact of Fragmentation, Exchange Fees and Liquidity Provision on Market Quality." *Journal of Empirical Finance* 41 (March): 140–60.

Antill, Samuel, and Darrell Duffie. Forthcoming. "Augmenting Markets with Mechanisms." *Review of Economic Studies*.

Baldauf, Markus, and Joshua Mollner. 2020. "High-Frequency Trading and Market Performance." *Journal of Finance* 75 (3): 1495–526.

Baldauf, Markus, and Joshua Mollner. 2021. "Trading in Fragmented Markets." *Journal of Financial and Quantitative Analysis* 56 (1): 93–121.

Bennett, Paul, and Li Wei. 2006. "Market Structure, Fragmentation, and Market Quality." *Journal of Financial Markets* 9 (1): 49–78.

Bernales, Alejandro, Nicolás Garrido, Satchit Sagade, Marcela Valenzuela, and Christian Westheide. 2018. "A Tale of One Exchange and Two Order Books: Effects of Fragmentation in the Absence of Competition." SAFE Working Paper 234.

Boehmer, Beatrice, and Ekkehart Boehmer. 2003. "Trading Your Neighbor's ETFs: Competition or Fragmentation?" *Journal of Banking and Finance* 27 (9): 1667–703.

Brown, Donald J., and Stephen A. Ross. 1991. "Spanning, Valuation and Options." *Economic Theory* 1 (1): 3–12.

Budish, Eric, Peter Cramton, and John Shim. 2015. "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." *Quarterly Journal of Economics* 130 (4): 1547–621.

Budish, Eric, Robin Lee, and John Shim. 2019. "Will the Market Fix the Market? A Theory of Stock Exchange Competition and Innovation." Unpublished.

Chakravarty, Sugato, Pankaj Jain, James Upson, and Robert Wood. 2012. "Clean Sweep: Informed Trading through Intermarket Sweep Orders." *Journal of Financial and Quantitative Analysis* 47 (2): 415–35.

Chao, Yong, Chen Yao, and Mao Ye. 2019. "Why Discrete Price Fragments U.S. Stock Exchanges and Disperses Their Fee Structures." *Review of Financial Studies* 32 (3): 1068–101.

Chlistalla, Michael, and Marco Lutat. 2011. "Competition in Securities Markets: The Impact on Liquidity." *Financial Markets and Portfolio Management* 25 (2): 149–72.

Chowdhry, Bhagwan, and Vikram Nanda. 1991. "Multimarket Trading and Market Liquidity." *Review of Financial Studies* 4 (3): 483–511.

Chung, Kee H., and Chairat Chuwonganant. 2012. "Regulation NMS and Market Quality." *Financial Management* 41 (2): 285–317.

Clapham, Benjamin, Peter Gomber, Jens Lausen, and Sven Panz. 2021. "Liquidity Provider Incentives in Fragmented Securities Markets." *Journal of Empirical Finance* 60: 16–38.

Colliard, Jean-Edouard, and Thierry Foucault. 2012. "Trading Fees and Efficiency in Limit Order Markets." *Review of Financial Studies* 25 (11): 3389–421.

De Fontnouvelle, Patrick, Raymond P. H. Fishe, and Jeffrey H. Harris. 2003. "The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999." *Journal of Finance* 58 (6): 2437–63.

Degryse, Hans, Frank de Jong, and Vincent van Kervel. 2015. "The Impact of Dark Trading and Visible Fragmentation on Market Quality." *Review of Finance* 19 (4): 1587–622.

Du, Songzi, and Haoxiang Zhu. 2017. "What Is the Optimal Trading Frequency in Financial Markets?" *Review of Economic Studies* 84 (4): 1606–51.

Duffie, Darrell, and Haoxiang Zhu. 2017. "Size Discovery." *Review of Financial Studies* 30 (4): 1095–150.

Félez-Viñas, Ester. 2017. "Market Fragmentation, Mini Flash Crashes and Liquidity." Unpublished.

Foley, Sean, Elivs Jarnecic, and Anqi Liu. 2020. "Forming an Orderly Line: How Queue-Jumping Drives Fragmentation." Unpublished.

Foucault, Thierry, and Albert J. Menkveld. 2008. "Competition for Order Flow and Smart Order Routing Systems." *Journal of Finance* 63 (1): 119–58.
Gentile, Monica, and Simone Fioravanti. 2011. "The Impact of Market Fragmentation on European Stock Exchanges." Unpublished.
Glosten, Lawrence R. 1994. "Is the Electronic Open Limit Order Book Inevitable?" *Journal of Finance* 49 (4): 1127–61.
Gomber, Peter, Satchit Sagade, Erik Theissen, Moritz Christian Weber, and Christian Westheide. 2016. "Spoilt for Choice: Order Routing Decisions in Fragmented Equity Markets." SAFE Working Paper 143.
Gomber, Peter, Satchit Sagade, Erik Theissen, Moritz Christian Weber, and Christian Westheide. 2017. "Competition between Equity Markets: A Review of the Consolidation versus Fragmentation Debate." *Journal of Economic Surveys* 31 (3): 792–814.
Gresse, Carole. 2012. *Market Fragmentation in Europe: Assessment and Prospects for Market Quality*. The Future of Computer Trading in Financial Markets: Foresight Driver Review 19. United Kingdom: Government Office for Science.
Gresse, Carole. 2017. "Effects of Lit and Dark Market Fragmentation on Liquidity." *Journal of Financial Markets* 35: 1–20.
Hamilton, James L. 1979. "Marketplace Fragmentation, Competition, and the Efficiency of the Stock Exchange." *Journal of Finance* 34 (1): 171–87.
Haslag, Peter, and Matthew Ringgenberg. 2020. "The Causal Impact of Market Fragmentation on Liquidity." Paper presented at the Fourth Annual Conference on Financial Market Regulation, Washington, DC.
Hengelbrock, Jödis, and Erik Theissen. 2009. "Fourteen at One Blow: The Market Entry of Turquoise." Unpublished.
Kawakami, Kei. 2017. "Welfare Consequences of Information Aggregation and Optimal Market Size." *American Economic Journal: Microeconomics* 9 (4): 303–23.
Klemperer, Paul D., and Margaret A. Meyer. 1989. "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica* 57 (6): 1243–77.
Körber, Lena, Oliver B Linton, and Michael Vogt. 2013. "The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market." Unpublished.
Malamud, Semyon, and Marzena Rostek. 2017. "Decentralized Exchange." *American Economic Review* 107 (11): 3320–62.
Malinova, Katya, and Andreas Park. 2019. "Sniping in Fragmented Markets." Unpublished.
Mendelson, Haim. 1987. "Consolidation, Fragmentation, and Market Performance." *Journal of Financial and Quantitative Analysis* 22 (2): 189–207.
Menkveld, Albert J. 2008. "Splitting Orders in Overlapping Markets: A Study of Cross-Listed Stocks." *Journal of Financial Intermediation* 17 (2): 145–74.
O'Hara, Maureen, and Mao Ye. 2011. "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics* 100 (3): 459–74.
Pagano, Marco. 1989. "Endogenous Market Thinness and Stock Price Volatility." *Review of Economic Studies* 56 (2): 269–87.
Pagnotta, Emiliano S., and Thomas Philippon. 2018. "Competing on Speed." *Econometrica* 86 (3): 1067–115.
Parlour, Christine A., and Duane J. Seppi. 2003. "Liquidity-Based Competition for Order Flow." *Review of Financial Studies* 16 (2): 301–43.
Rostek, Marzena, and Ji Hee Yoon. 2020. "Exchange Design and Efficiency." Unpublished.
Rostek, Marzena, and Marek Weretka. 2012. "Price Inference in Small Markets." *Econometrica* 80 (2): 687–711.
Rostek, Marzena, and Marek Weretka. 2015. "Dynamic Thin Markets." *Review of Financial Studies* 28 (10): 2946–92.
Rust, John, and George Hall. 2003. "Middlemen versus Market Makers: A Theory of Competitive Exchange." *Journal of Political Economy* 111 (2): 353–403.
Sannikov, Yuliy, and Andrzej Skrzypacz. 2016. "Dynamic Trading: Price Inertia and Front-Running." Unpublished.
Spankowski, Ulli, Martin Wagener, and Hans-Peter Burghof. 2012. "The Role of Traditional Exchanges in Fragmented Markets." Unpublished.
Stoll, Hans R. 2001. "Market Fragmentation." *Financial Analysts Journal* 57 (4): 16–20.
van Kervel, Vincent, Amy Kwan, and Joakim Westerholm. 2020. "Order Splitting and Interacting with a Counterparty." Unpublished.
Vayanos, Dimitri. 1999. "Strategic Trading and Welfare in a Dynamic Market." *Review of Economic Studies* 66 (2): 219–54.

Vives, Xavier. 2011. "Strategic Supply Function Competition with Private Information." *Econometrica* 79 (6): 1919–66.
Wilson, Robert. 1979. "Auctions of Shares." *Quarterly Journal of Economics* 93 (4): 675–89.
Wittwer, Milena. 2021. "Connecting Disconnected Financial Markets?" *American Economic Journal: Microeconomics* 13 (1): 252–82.
Zhu, Haoxiang. 2014. "Do Dark Pools Harm Price Discovery?" *Review of Financial Studies* 27 (3): 747–89.

This article has been cited by:

1. Justin Cox, Bonnie Van Ness, Robert Van Ness. 2022. The dark side of IPOs: Examining where and who trades in the IPO secondary market. *Financial Management* **51**:4, 1091-1126. [Crossref]
2. Guangcheng Ma, Jianhua Cao. 2022. Spatial Heterogeneity Impacts of Bilateral Foreign Direct Investment on Green Energy Efficiency in China. *Frontiers in Environmental Science* **10**. . [Crossref]
3. Wen Guo, Xiaorui Liu. 2022. Market fragmentation of energy resource prices and green total factor energy efficiency in China. *Resources Policy* **76**, 102580. [Crossref]
4. Laurence Daures-Lescourret, Sophie Moinas. 2022. Fragmentation and Strategic Market-Making. *Journal of Financial and Quantitative Analysis* 1-26. [Crossref]
5. Suchismita Mishra, Le Zhao. 2021. Order Routing Decisions for a Fragmented Market: A Review. *Journal of Risk and Financial Management* **14**:11, 556. [Crossref]
6. Fabricius Somogyi. 2021. Dollar Dominance in FX Trading. *SSRN Electronic Journal* **5**. . [Crossref]
7. Marzena J. Rostek, Ji Hee Yoon. 2020. Innovation in Decentralized Markets. *SSRN Electronic Journal* **53**. . [Crossref]
8. Jingxiong Hu, Canyao Liu, Jiaheng Yu. 2020. Can Off-Exchange Trading Improve Close Price Informativeness?. *SSRN Electronic Journal* **74**. . [Crossref]
9. Tom Grimstvedt Meling, Bernt Arne degaard. 2017. Tick Size Wars, High Frequency Trading, and Market Quality. *SSRN Electronic Journal* **41**. . [Crossref]
10. Tomy Lee, Chaojun Wang. 2017. Why Trade Over-the-Counter? When Investors Want Price Discrimination. *SSRN Electronic Journal* **4**. . [Crossref]
11. Laurence Lescourret, Sophie Moinas. 2014. Liquidity Supply Across Multiple Trading Venues. *SSRN Electronic Journal* **56**. . [Crossref]