The Persistence of Market Fragmentation and Liquidity

Michael Coccia[†*]

This Draft: September 11, 2026

For the most recent version click here

Abstract

More than half of U.S. equity volume executes away from exchanges that display prices, and the remainder is spread across seventeen exchanges. Whether this dispersion of trading deteriorates market quality is disputed in theory and unresolved in the data. Using a decade of high-frequency data for 1,758 highly traded stocks, I document that both the allocation of a stock's trading across venues and its liquidity are long memory processes with common components. Previous econometric specifications do not accurately identify the relation between fragmentation and liquidity given these properties. I find that competition among lit exchanges is associated with tighter spreads at short horizons and wider spreads at longer horizons, with deeper books at every horizon. A stock's order flow migration off-exchange is associated with wider lit spreads and thinner lit depth at every horizon, while market-wide off-exchange migration is associated with deeper books.

JEL Codes: G14, G18, C22, C23
Keywords: Market Fragmentation, Long Memory, Fractional Integration, Liquidity, Market Microstructure, Common Factor, Frequency Analysis

[*]Department of Finance, University of Mississippi. Email: macoccia@olemiss.edu. I would like to thank my advisor Dr. Robert Van Ness for helpful comments. All errors are my own.

[†] I acknowledge the Mississippi Center for Supercomputing Research (MCSR) at the University of Mississippi for providing the computing resources and technical support that contributed to the research results reported within this paper

1 Introduction

Fragmentation of trading activity across U.S. exchanges is now a fundamental component of modern equity markets, but this was not always the case. The shift is largely a byproduct of Regulation National Market System (Reg NMS), which, when passed by the Securities and Exchange Commission (SEC) in 2005, completely rewired U.S. equity markets. The goal of Reg NMS was to modernize and strengthen U.S. equity trading. To communicate the magnitude of the shift, consider the following statistics. In January 2005, before the rule took effect, the NYSE executed 79.1% of the volume in the stocks it listed with most of the remainder filled in-house by broker-dealers.[1] By October 2009 the NYSE's share of volume in its listed stocks had fallen to 25.1%, while across all NMS stocks approximately 32 dark pools executed 7.9% of share volume and more than 200 internalizing broker-dealers another 17.5%.[2] By 2022 sixteen exchanges executed 59.7% of volume,[3] and in September 2026 the trade reporting facilities, which collect and aggregate off-exchange trades, reported 53% of the consolidated volume.[4] Figure 1 shows how volume is spread across venues in May 2026. Half of the volume prints off-exchange, the Nasdaq and the NYSE together constitute about 25%, and ten of the seventeen exchanges each execute less than one percent. This fragmentation of activity in equity markets has become a primary concern of regulators. The Chairman of the SEC delivered the verdict on Reg NMS at a December 2025 roundtable: *"Two decades have given us the benefit of perspective, and the verdict is clear. Reg NMS, built on flawed foundations, has invited gamesmanship and contributed to the fragmentation of our markets, to the dispersal of liquidity, and diminished transparency"* (U.S. Securities and Exchange Commission, 2025).

This splintering of activity across visible stock exchanges and off-exchange venues has drawn the attention of not only regulators, but has also inspired an extensive empirical literature. The results of that literature are inconsistent. Some find fragmentation harmful, some find it beneficial, and some studies present conflicting results within their own analysis and suggest that the relation must be sample dependent. In this study, I document a new statistical property of fragmentation motivated by theory and exploit the new property to explain the conflicting findings in the literature.

Trading venues differ in what they display, whom they serve and what they cost to trade. A national securities exchange displays its best bid and offer in the consolidated feed and is often referred to as a 'lit' exchange. Off-exchange trading is primarily composed of but not limited to: alternative

[1] See (U.S. Securities and Exchange Commission, 2010; Staff of the Division of Trading and Markets, 2013)

[2] See (U.S. Securities and Exchange Commission, 2010)

[3] See (U.S. Securities and Exchange Commission, 2022a)

[4] See (Cboe Global Markets, 2026)

trading systems, or dark pools, which are exchange-like matching engines that do not display quotations in the consolidated feed; wholesalers, which are off-exchange market makers that solicit the orders of retail brokers through payment for order flow and fill those orders against their own inventory; and single-dealer platforms, which are electronic venues with one dealer on the other side of every trade, and which primarily seek institutional orders for execution against that dealer's own inventory (U.S. Securities and Exchange Commission, 2022a). Off-exchange venues are often referred to as 'dark' venues since they do not display their quotes in the consolidated feed, but rather they use the quotes on lit venues to arrange trades.

The two venue types compete for order flow on different margins. Lit exchanges compete with one another through the display of their quotes, and while this competition channel can lower trading costs, it also exposes the liquidity providers to adverse selection and to their quotes being "sniped" by faster traders (Baldauf and Mollner, 2021; Aquilina, Budish and O'Neill, 2022). A snipe occurs when new information arrives and a faster trader executes against a liquidity provider's stale quote before that provider can cancel it (Budish, Cramton and Shim, 2015). Off-exchange venues compete differently than the lit exchanges. Wholesalers, for example, "generally attract order flow by offering to on average execute orders at prices that are better than displayed prices," and they "bundle their market access services with execution services, thereby vertically fully integrating order handling and execution services for their retail broker customers" (U.S. Securities and Exchange Commission, 2022a). The negative effect of off-exchange fragmentation is that potentially 'uninformed' order flow leaves the lit exchange, and faced with a higher proportion of 'informed' traders, market makers widen their quotes in response, deteriorating market quality (Glosten and Milgrom, 1985). Both types of fragmentation have the potential to positively or negatively influence market quality, and empirical studies typically test which one dominates. For example, Gresse (2017) frames her own study as setting out "to test which of the positive and negative effects of dark trading on liquidity dominates."

Within this fragmented setting, all market incumbents have an interest in how the allocation of order flow across venues affects market quality. For example, a retail investor's broker routes more than 90% of that investor's marketable orders to a wholesaler (U.S. Securities and Exchange Commission, 2022a,b), who fills them at a quote based off of the displayed national best bid or offer (NBBO). The wholesaler is not required to give any other market participant an opportunity to compete for the order, nor is the retail trader able to observe the price of a competitive fill (U.S. Securities and Exchange Commission, 2022a). The retail trader's welfare therefore depends on lit fragmentation, because the trade price is pegged to the displayed quote, and on how dark fragmentation influences that quote. An institutional trader may slice a parent order into child orders across seventeen exchanges (NYSE, 2025) and the thirty-plus dark pools that trade NMS stocks (U.S. Securities and Exchange Commission, 2022a).[5] The institutional trader's welfare depends on the price and liquidity available within the fragmented market, and liquidity becomes increasingly invisible as dark fragmentation increases. A liquidity provider must determine how much to display on each book and at what size, while balancing the risk of being picked off by another trader. A broker-dealer must decide whether to connect to every exchange or potentially route through someone that does; in the second quarter of 2025, of 223 broker-dealers that executed on an exchange, only 22 executed on all sixteen and 88 executed on a single one (Staff of the Office of Analytics and Research, Division of Trading and Markets, 2025). And finally, the SEC must decide whether a tick, fee cap or a trade-through rule written for the market of 2005 still best serves the modern market. All of these points circle back to one fundamental question. Does spreading a stock's trading across venues improve or deteriorate market quality, and does it depend on which type of fragmentation?

This paper measures the association between the dispersion of trading activity and market quality, allowing for this association to differ along two dimensions, whereas previous literature has forced the relation into a single point coefficient. The first dimension is the horizon over which the change in venue shares and market quality are analyzed (frequency), and the second is whether the change observed is attributable to a stock's own change or related to overall market trends (component). By allowing, but not forcing, the relation between fragmentation and liquidity to differ across the frequency $\times$ component spectrum I find evidence of both the positive and negative channels of fragmentation. Competition among lit exchanges is associated with tighter spreads at short horizons and wider spreads at long horizons, and with deeper books at every horizon. A stock's own migration off-exchange is associated with wider lit spreads and thinner lit depth at every horizon, whereas a market-wide migration off-exchange is associated with deeper books. I discuss how several papers with seemingly different conclusions or findings can be in agreement when placed within my findings.

The case for lit exchange competition is ordinary economics where competition lowers prices, and the SEC has described its duty since Reg NMS as facilitating "an appropriately balanced market structure that promotes competition among markets, while minimizing the potentially adverse effects of fragmentation" (U.S. Securities and Exchange Commission, 2010; Staff of the Division of Trading and Markets, 2013). Where the market sits on that balance is an empirical question, and decades of

[5] Sixteen exchanges traded NMS stocks from 2020 through 2024, and thirty-two alternative trading systems executed 10.2% of NMS stock volume in the first quarter of 2022 (U.S. Securities and Exchange Commission, 2022a). The seventeenth exchange is 24X National Exchange, which the Daily TAQ specification adds as participant ID "G" in version 4.1b, dated 7 August 2025 (NYSE, 2025). A stock in my sample prints on at most seventeen exchanges in 2025 and 2026 and on at most sixteen in 2022 through 2024. TXSE, participant ID "F", was added in version 4.3 and was scheduled to begin trading on 1 June 2026; it prints no volume in my sample (NYSE, 2026).

evidence have not settled it. The findings are heterogeneous enough that Baldauf and Mollner (2021) sort them into three groups. The first group reports a positive association between fragmentation and liquidity (Battalio, 1997; Fink, Fink and Weston, 2006; Nguyen, Van Ness and Van Ness, 2007; Foucault and Menkveld, 2008), the second reports a negative association (Arnold et al., 1999; Hendershott and Jones, 2005; Bennett and Wei, 2006), and the third reports an inverted-U relation in which liquidity is highest at moderate fragmentation (Degryse, De Jong and Kervel, 2015). Their own estimation falls in the second group. I would add to this list the studies that find the relation to be conditional on where a stock sits in the cross-section (Gresse, 2017; Haslag and Ringgenberg, 2023). Even the relation to seemingly similar mechanisms can produce different results. Consider the exposure channel which exists due to fast algorithmic traders and a continuous-time market structure (Budish, Cramton and Shim, 2015; Baldauf and Mollner, 2021). Yet Hendershott, Jones and Menkveld (2011) find that when the NYSE automated its quotes the resulting rise in algorithmic trading narrowed spreads for large stocks, with no significant effect for smaller ones, whereas Aquilina, Budish and O'Neill (2022) examine message data from the London Stock Exchange and find that latency arbitrage races between algorithms account for a fifth of volume and one-third of price impact. Recent U.S. evidence reports that large stocks' market quality has benefited relative to small stocks' in relation to fragmentation (Haslag and Ringgenberg, 2023).

While lit fragmentation's effect is generally described as a tradeoff between the competition and exposure channels, off-exchange fragmentation is typically predicted to harm lit exchange liquidity, by two different routes. In Zhu (2014) uninformed traders self-select into the dark pool, so that "the addition of a dark pool tends to increase order informativeness, spreads, and price impacts on exchanges." In Buti, Rindi and Werner (2017) there are no informed traders at all and instead "order migration away from lit markets to dark pools may adversely influence the incentive for traders to provide liquidity in the lit market, potentially resulting in higher trading costs." Comerton-Forde and Putniņš (2015) find evidence of partial segmentation deteriorating price discovery in the Australian markets, though only above a threshold level of dark trading. Foley and Putniņš (2016) find that two-sided dark trading benefits market quality while the effects of one-sided dark trading are weaker and inconclusive. Buti, Rindi and Werner (2022) state that the relation between dark trading and market quality depends on the form of dark trading, the part of the cross-section and the time period observed.

Three seminal papers discuss the concentration of traders, where concentration can be thought of as the opposite of fragmentation. These papers do not mention the statistical property long memory, and none is a dynamic model, but each implies something about what holds an allocation of trading in place. I use the statistical property of long memory in time series data to ask which account the data appears to represent. In Pagano (1989) each trader locates depth by conjecturing where others will locate it and the conjectures may become self-fulfilling. Therefore, an allocation survives because everyone expects it to, and only "a sufficiently dramatic change in conjectures" empties one market into the other. In contrast, Chowdhry and Nanda (1991) show that discretionary traders gather in the market that holds the largest number of traders who cannot move between venues, so the allocation is inherited from the distribution of those immobile traders rather than produced by coordination. Yet in the setting of Glosten (1994), where entry is costless and orders split across venues at no cost, the allocation does not matter at all. From these three papers, concentration is stable until beliefs change in the first, inherited from immobile order flow in the second and irrelevant in the third.

Long memory is a statistical property in time series data which means that shocks decay much more slowly than they do under a short-memory process. Long memory can otherwise be referred to as 'persistence' in the data. Long memory also implies that variation in the data accumulates at lower frequencies (longer periods) (Baillie, 1996). A single OLS coefficient is a variance-weighted average of relations that hold at each frequency, and under long memory that variance sits almost entirely at low frequencies (Engle, 1974). The coefficient then reports how the slow-moving components of the two series line up, rather than the day-to-day relation it is usually interpreted as. The asset pricing literature has outlined this concern. Bansal and Yaron (2004) show that a small persistent component in consumption growth, one that explains little of the variation in monthly data, carries a large risk premium precisely because the price of that risk rises with its persistence. Bryzgalova, Huang and Julliard (2025) represent a macro factor as a moving average of priced asset-return shocks and find that macro risk premia are negligible in the short run yet match the equity premium at business cycle horizons, and they read this horizon dependence as "reconciling the seemingly discordant findings across frequencies in the previous literature." The subject and the methodology differ from this paper, but the takeaway is the same.[6] A relation estimated at one horizon need not describe the relation at another. The fragmentation literature has also hinted at the existence of a horizon-dependent channel as Zhu (2014) makes the horizon of information a parameter of the model and Buti, Rindi and Werner (2017) note that the effects of introducing a dark pool can vary over time. No empirical paper has examined the relation of fragmentation and market quality across varying time horizons.[7] I investigate the relation of fragmentation and liquidity across two dimensions, frequency and component. Long

[6]Their design recovers the horizon dependence using an estimated propagation of identified shocks. My panel of venue shares and spreads does not provide me either the pricing restrictions or the identified shocks. Further, a moving average truncated at a few dozen lags cannot represent innovations which decay hyperbolically. This is why I use the fast Fourier transform (FFT) on my observed sample and estimate the relation on the frequency level.

[7]Menkveld, Yueshen and Zhu (2017) do model a dynamic relation of liquidity and off-exchange venues using a VARX model. Their time horizon is intraday whereas my shortest time horizon is two days and the longest is 500+ days.

memory motivates the inspection across the frequency dimension.

I also contend that it is reasonable to investigate not only how the variation of fragmentation within a stock is related to liquidity, but also the variation in a stock's fragmentation as a component of overall market shifts in fragmentation. The determinants of fragmentation in an individual stock may be a byproduct of regulations, in addition to market conditions which affect all stocks. Consider Reg NMS, which set the order protection rule and the access fee cap for every listed stock at once (U.S. Securities and Exchange Commission, 2005). It may be the case that systematic and idiosyncratic channels that influence a stock's fragmentation yield different relations to the same stock's liquidity. Further motivation is that liquidity itself has been documented with a common component. The common component was first identified in Chordia, Roll and Subrahmanyam (2000), Hasbrouck and Seppi (2001) and Huberman and Halka (2001), who document by three different designs that individual stocks' spreads and depths co-move with a common liquidity factor, though they disagree on how much of the variation it explains. Later work on the commonality of liquidity examines price exposure to the common factor (Korajczyk and Sadka, 2008), the time series progression of commonality in liquidity (Kamara, Lou and Sadka, 2008), and attributes the mechanism to correlated trading by mutual funds (Koch, Ruenzi and Starks, 2016). Aggregate liquidity also has been documented to predict real macroeconomic activity (Chen, Eaton and Paye, 2018). This motivates the second dimension of the frequency $\times$ component spectrum I analyze. The distinction of a common factor also pertains to how the literature has previously identified estimates. A common design has been to instrument a stock's variable with the average of said variable of other stocks in its size group or other grouping condition (Hasbrouck and Saar, 2013; Comerton-Forde and Putniņš, 2015; Degryse, De Jong and Kervel, 2015; Buti, Rindi and Werner, 2022). The instrument is adopted to break reverse causality, and the relevance is argued from commonality as Degryse, De Jong and Kervel (2015, p. 1607) write that "the instruments create variation in fragmentation (and dark trading) stemming from the commonality of fragmentation (and dark trading) across stocks that is orthogonal to the commonality in liquidity across stocks." The identifying variation is therefore the common component by construction. I show that this changes which relation is estimated, because the common and idiosyncratic components of fragmentation often carry different associations with liquidity.

Not only is this paper connected to the literature on lit and dark market fragmentation and commonality in liquidity, but it is also linked to the literature on long memory processes in finance data. Long memory in a financial time series was famously established for volatility by Ding, Granger and Engle (1993) and spawned countless (generalized) autoregressive conditional heteroskedasticity models (GARCH). Long memory has since been documented in volume (Bollerslev and Jubinski, 1999;

Lobato and Velasco, 2000), intraday order flow direction (Lillo and Farmer, 2004; Bouchaud, Farmer and Lillo, 2009), and aggregate liquidity (Chen, Eaton and Paye, 2018). My study adds daily stock-level fragmentation, daily aggregate fragmentation and daily stock-level liquidity to the family of long memory time series in financial data. I exploit this property within fragmentation and liquidity to provide an explanation for the conflicting existing literature.

The final strand of literature my paper connects to is frequency analysis, often referred to as spectral analysis. An early example, whose methodology this paper draws on, is Engle (1974) who studies Friedman's permanent income hypothesis by Fourier transforming the data and subsequently running OLS on the "bands" of frequencies.[8] Spectral analysis has been used in the asset pricing literature to identify the cyclical components of the consumption growth time series (Ortu, Tamoni and Tebaldi, 2013) and to quantify the price risk of consumption fluctuations at each frequency (Dew-Becker and Giglio, 2016). The literature on spectral analysis demonstrates that heterogeneous horizon-specific relations across frequencies can be hidden by aggregating data (Bandi et al., 2021; Bandi and Tamoni, 2023). I analogously investigate the relation of fragmentation and liquidity across the frequency spectrum, without forcing heterogeneity and find the relations differ in magnitude and direction across the spectrum. Frequency analysis and Fourier transformations are also implemented in microstructure settings, although the timescale is usually intraday (Sato, 2008; Malliavin and Mancino, 2009; Gradojevic, 2012; Bibinger and Reiß, 2014; Gould, Porter and Howison, 2016; Firouzi and Wang, 2019; Zhao et al., 2025; Wu, Zhang and Dai, 2026).

My empirical analysis contains 1,758 common stocks from January 2016 through June 2026, roughly 3.8 million stock-days. From the NYSE Trade and Quote data I compute the degree of both lit and dark fragmentation. Specifically, I construct a Herfindahl-Hirschman index of exchange venue shares to measure lit fragmentation and use the proportion of volume executed off-exchange to measure dark fragmentation. I compute the effective spread to measure trading costs and the depth at the NBBO to measure liquidity. I use on-exchange trades and quotes only, ordering every trade and quote by the time stamp of the exchange that produced it rather than by the consolidated tape, as Battalio et al. (2026) show that the tape's ordering mis-signs trades and understates effective spreads by over 40% for trades with high exchange-to-SIP latency. My sample excludes thinly traded stocks through a firm-level activity screen, but covers about 87% of the dollar volume and 89% of the market capitalization of all common stocks on the average day. The stocks in the sample have a median market capitalization of about $6 billion and the tenth percentile of market cap is near $1 billion. In my

[8] Fourier transformations have been commonly used in asset pricing or option pricing models, for example see Carr and Madan (1999); Duffie, Pan and Singleton (2000); Singleton (2001); Carr and Wu (2004); Eraker (2004); Aït-Sahalia and Kimmel (2007).

sample the equal-weighted off-exchange share rises from about 30% in 2016 to near 50% in 2026.

I first outline the estimator I use to detect long memory. I then document the long memory property on the stock-level series of fragmentation and market quality. In the appendix I present evidence that the estimations on the stock-level are robust to detrending, tapering and the breaks test of Qu (2011). I test whether there is cointegration between the series and find no evidence of cointegration which rules out the use of a fractionally cointegrated vector error correction model (FCVECM) or a fractionally cointegrated vector autoregression model (FCVAR). I then display that the persistence differs in the cross-section. Following this I estimate panel regressions using a two-way fixed effects specification, a panel specification which removes persistence and common effects (Ergemen and Velasco, 2017), and an instrumental variables (IV) specification. Lastly I decompose the relations between fragmentation and market quality across the frequency and component spectra. I then discuss the implications of my results and the integration of existing literature into my findings.

The extant literature on market fragmentation almost exclusively forces the relation between a type of fragmentation and a dimension of market quality into a single point coefficient. I find that in a classic two-way fixed-effects panel: both lit and dark fragmentation are associated with wider spreads, dark fragmentation is negatively related to depth, and lit fragmentation is positively related to depth. After I make the data stationary and remove the common component, the magnitudes shift and the association between lit fragmentation and trading costs flips. When estimating the relations using the IV approach commonly used in the literature (Hasbrouck and Saar, 2013; Comerton-Forde and Putniņš, 2015; Degryse, De Jong and Kervel, 2015; Buti, Rindi and Werner, 2022), I find that three of the four associations flip in sign relative to the levels specification. The disagreement in the relations, and an explanation of what drives the differences are key results of the paper.

When I permit the associations between the time series to differ across horizons, the lit association with trading costs flips at a 125-day period. Below 125 days, an increase in fragmentation tightens spreads, but above 125 days an increase in fragmentation widens spreads. Across all time periods, the relation of the common and idiosyncratic components of fragmentation and spreads share the same sign. For dark fragmentation, I find that the relation with respect to trading costs is near zero or positive at all period lengths and with respect to depth it is negative for periods under 250 days and positive at periods above 500 days. The associations of dark fragmentation and market quality are complex once the relation is decomposed into the common and idiosyncratic components: the idiosyncratic relation of dark fragmentation and trading costs is positive at all frequencies, the common relation of dark fragmentation and trading costs is negative below 40 days and positive above 40 days, the idiosyncratic relation between dark fragmentation and depth is negative for all periods, and the common relation between dark fragmentation and depth is positive at all frequencies. I interpret all of my estimates only as associations. The design includes stock fixed effects, heterogeneous loadings on persistent common factors and memory removal from both components, but it does not identify causal effects.

Results which vary over which frequency and component are measured need not be surprising given the fragmentation literature is yet to reach a consensus of the effects on market quality. In the final section of the paper I present a mapping of existing theoretical predictions and empirical results to the frequency $\times$ component spectrum documented in the paper.

The rest of my paper is organized as follows: Section 2 reviews related literature; Section 3 describes my data source, sample construction and definitions of fragmentation and liquidity variables; Section 4 details the identification of long memory in fragmentation and related time series, and examines cointegration between the series; Section 5 examines cross-sectional variation in persistence; Section 6 examines the relation between fragmentation and liquidity, via OLS with two-way fixed effects, the Ergemen and Velasco (2017) methodology for fractionally integrated panels with fixed effects and cross-sectional dependence, the Degryse, De Jong and Kervel (2015) 2SLS specification and an examination across the frequency $\times$ component spectrum; Section 7 discusses and relates my findings relative to extant literature; Section 8 concludes.

2 Related Literature

2.1 Long Memory in Financial Time Series

Baillie (1996) provides a comprehensive survey of long memory processes and fractional integration in econometrics and discusses the many ways in which a long memory process can be defined. A long memory process can be thought of most simply as a time series in which the autocorrelations decay hyperbolically (k^{2d-1}) at lag k rather than exponentially (ρ^k); equivalently, in the frequency domain, as one whose spectral density diverges at the origin with the rate λ^{-2d}, and d being the differencing parameter. This implies that shocks to a process with long memory decay more slowly than those of a short memory process such as a stationary autoregressive moving average (ARMA) process. Such a process can be modeled as an autoregressive fractionally integrated moving average (ARFIMA) process with parameters (p, d, q), where p and q are the autoregressive and moving average orders respectively, and d is the fractional integration/differencing parameter.

Perhaps the first application of the R/S statistic to stock returns is Greene and Fielitz (1977), who find evidence of long-term dependence in 200 daily NYSE return series. Lo (1991) shows that the statistic they use is biased toward finding dependence when short-range dependence is present, develops the modified R/S statistic which is robust to it, and finds no evidence of long memory in stock returns once that correction is applied. Ding, Granger and Engle (1993) find that absolute returns and squared returns (along with returns to other powers) exhibit long memory, while returns themselves do not.[9] Bollerslev and Jubinski (1999) extend the long memory literature to trading volume under a mixture of distributions hypothesis (MDH) in which the latent information-arrival process is fractionally integrated. The MDH implies absolute returns and volume share a common order of integration, and they estimate a median $\hat{d}$ of 0.40 for both absolute returns and turnover.[10] Lobato and Velasco (2000) revisit this result for log trading volume using an estimator which remains valid when the memory parameter lies outside the stationary region, and find that most estimates lie between 0.3 and 0.5, with some above 0.5. While Lobato and Velasco (2000) document a shared long memory parameter in both volume and volatility, they reject the two series sharing a long memory component, implying they are not fractionally cointegrated.

The persistence of an intraday time series can be traced to Andersen and Bollerslev (1997) who find evidence that numerous heterogeneous short-run information arrivals can manifest themselves as an aggregate long memory process in volatility. The persistence of intraday order flow was identified before it was classified as long memory with Hasbrouck (1991) who models trades and quote revisions as a vector autoregression (VAR) and reports strong positive autocorrelations in trades. Hasbrouck (1991) further notes that order flow is affected by prior quote revisions and that the full price impact of a trade arrives only with a protracted lag. Lillo and Farmer (2004) directly classify order flow as long memory and show that the signs (buy or sell) of orders on the London Stock Exchange follow a long memory process.[11] Bouchaud, Farmer and Lillo (2009) corroborate the long memory finding in order flow by investigating the London, Paris, New York and Spanish markets and posit that since displayed liquidity pales in comparison to institutional order flow demand, the splitting of orders over days or months results in the observed long memory process. Chen, Eaton and Paye (2018) extend long memory to aggregate illiquidity, estimating $\hat{d}$ between 0.35 and 0.89 across a range of illiquidity proxies using the same ELW estimator I employ in this paper. Long memory has thus been documented in

[9] Many studies have since investigated long memory in volatility and proposed alternate estimating and modeling techniques, for example see Breidt, Crato and De Lima (1998); Comte and Renault (1998); Hurvich, Moulines and Soulier (2005); Harvey (2007); Corsi (2009). This list is not exhaustive.

[10] See Clark (1973); Epps and Epps (1976); Tauchen and Pitts (1983) for the earlier theoretical work on MDH and predictions for the relation between volume and volatility.

[11] Lillo and Farmer (2004) note that the strong predictability in order flow they document would be a violation of market efficiency. The authors further demonstrate that the predictability in order flow is offset by compensating fluctuations in liquidity and market order size which themselves are a long memory process and the two counterbalance one another to leave directional price impact approximately uncorrelated.

stock volume, volatility, the direction of order flow, and the cost to trade. It has not been documented in the dispersion of where trading occurs.

2.2 Market Fragmentation and Venue Competition

Theory predicts that trading concentrates and disagrees about the persistence of said concentration. Pagano (1989) shows that traders cluster by the size of their desired transactions, and that concentration is generally Pareto-superior to two-market fragmentation.[12] Each trader acts on their own conjectures of where to locate depth, and these conjectures are self-fulfilling. Pagano further states that large shifts can cause agents to revise their beliefs and relocate all traders to one venue.[13] Admati and Pfleiderer (1988) derive the same coordination of traders in an intraday temporal setting rather than across venues, where liquidity traders cluster in periods that are already thick. Chowdhry and Nanda (1991) note that this result requires liquidity traders to be unable to split their trades across periods. Once order-splitting is allowed, concentration survives for a different reason: "the order flows from earlier periods are informative in later periods" (Chowdhry and Nanda, 1991, p. 494).[14] Concentration then depends on the history of order flow rather than on coordination alone.[15] In their own model, Chowdhry and Nanda find that discretionary liquidity traders concentrate in the market holding the largest number of traders who cannot move between venues. Persistence in their account is inherited from the distribution of those traders rather than produced by the allocation itself.

Glosten (1994) questions whether venue allocation is relevant. He demonstrates in a frictionless setting that aggregate liquidity is invariant to how trading is divided across venues and the limit order book is the only stable institution in his setting. In a setting closer to modern markets, Baldauf and Mollner (2021) return to the relevance of venue allocation by introducing frictions that prevent investors from always trading at the best available price.[16] They propose that fragmentation creates two opposing channels. On one hand, fragmentation increases competition and lowers trading costs but on the other hand, market makers are exposed to adverse selection in more fragmented markets. Which channel dominates is an empirical question rather than a theoretical one. They find that the exposure channel dominates, with a counterfactual monopoly spread 23% lower than the prevailing

[12] This conclusion depends on whether large traders are able to find better execution off-exchange with a search.

[13] This can be thought of as a "regime shift" switch rather than a stochastic trend, which makes the structural-break alternative to long memory a prediction of the theory rather than a purely statistical concern. I return to this topic in Appendix A.

[14] Lillo and Farmer (2004) document that order flow is in fact long memory, and Bouchaud, Farmer and Lillo (2009) attribute the persistence to order splitting.

[15] Chowdhry and Nanda (1991) further note that their own simultaneous-trading setting forecloses this channel.

[16] The friction is modeled as investor heterogeneity in the propensity to trade at the best available price, motivated by the cost of monitoring quotes in real time and by the agency problem between an investor and the broker routing the order.

duopoly spread.[17] Long memory can be measured by the differencing parameter d and none of the models discussed deliver a value of d. Nor do the models explicitly predict long memory in venue concentration, although the presence of long memory in fragmentation is a test of their predictions. Concentration is permanent in Pagano (1989), inherited from immobile order flow in Chowdhry and Nanda (1991), and irrelevant in Glosten (1994). Estimating d locates the persistence of fragmentation among these models rather than testing for direct acceptance or refutation of any specific model.

Hasbrouck (1995) describes fragmentation as "the dispersal of trading in a security to multiple sites" and as "a dominant institutional trend" three decades ago. He asks where price discovery occurs across those sites rather than what the dispersal does to market quality. Empirical work has since measured the relation of fragmentation with market quality, and it is extensive.[18] O'Hara and Ye (2011) provide one of the first market-level empirical investigations of fragmentation in modern U.S. equities using a venue share variable. They measure market fragmentation with the share of trading volume reported to trade reporting facilities (TRFs), and find that fragmentation is associated with lower trading costs and faster execution speeds.[19] Investigating the Dutch markets, Degryse, De Jong and Kervel (2015) build a Herfindahl-Hirschman index (HHI, the sum of the squared venue shares) to measure lit fragmentation and use the share of dark trading to measure dark fragmentation. Degryse et al. find that lit fragmentation is associated with increased global depth while dark trading is associated with deteriorated global depth and spreads.[20] Menkveld, Yueshen and Zhu (2017) provide one of the literature's few dynamic treatments of venue shares, estimating a panel vector autoregression with exogenous variables (VARX) on minute-level market shares to document a pecking order of venue types ordered by immediacy. Haslag and Ringgenberg (2023) examine the relation between fragmentation and market quality in U.S. equities from 2003 through 2016, measuring fragmentation as one minus the HHI of trade volume across the reporting venues in TAQ, which counts the trade reporting facility as a single venue and so does not separate lit from dark. They report that fragmentation improves market quality for large stocks while small stocks fare relatively worse. In each of these studies venue concentration appears on the right hand side as an explanatory variable for market quality, and where it is treated dynamically the horizon is intraday. I am aware of no study which asks whether the series itself carries memory at longer horizons or a study which considers the long memory of fragmentation

[17] The data used in their empirical test only contain one ETF in the Australian market.

[18] I discuss papers whose fragmentation measurement variables or empirical setting are most similar to mine, or employ a similar panel setting regression methodology. There is extensive literature which resides under the category of market fragmentation but may include cross-listings, liquidity provider entries, market-linkings etc. which may bundle additional effects which I do not aim to study.

[19] The results are found to differ across stock size, and whether a stock is listed on the NYSE or Nasdaq.

[20] While their main results document the positive association between lit fragmentation and global depth, they also document an inverted U-shape relation between fragmentation and global depth.

and liquidity when estimating relations.

3 Data and Measurement

3.1 Venue Concentration

I use the New York Stock Exchange (NYSE) Trade and Quote (TAQ) data to construct an HHI of exchange venue trading concentration. Let $V_{i,t,j}$ denote the share volume in stock i on day t at exchange j, let K denote the set of TAQ reporting venues, and let L denote on-exchange (lit) venues.

$$s_{i,t,j}^{Lit} = \frac{V_{i,t,j}}{\sum_{j \in L} V_{i,t,j}}, \qquad s_{i,t,j}^{All} = \frac{V_{i,t,j}}{\sum_{j \in K} V_{i,t,j}}, \tag{1}$$

I then use the venue share measures to compute the HHI as follows,

$$HHI_{i,t}^{Lit} = \sum_{k \in L} (s_{i,t,k}^{Lit})^2, \qquad HHI_{i,t}^{All} = \sum_{k \in K} (s_{i,t,k}^{All})^2, \tag{2}$$

where $HHI_{i,t}^{Lit}$ and $HHI_{i,t}^{All}$ measure the lit- and all-venue concentration of trading in stock i on day t. In recent work Haslag and Ringgenberg (2023) use $HHI_{i,t}^{All}$ as their proxy for fragmentation. The authors note that exchange "D" is an aggregation of trades which occur on many dark pools, ATS or are internalized and $HHI_{i,t}^{All}$ understates the true degree of total market fragmentation. I demonstrate in Table 6 that $HHI_{i,t}^{All}$ bundles both dark and lit fragmentation, which often have opposite relations, into one coefficient. I therefore detail the construction of both here, but will rely on $HHI_{i,t}^{Lit}$ in my subsequent estimations of the association between lit fragmentation and liquidity. Following Haslag and Ringgenberg (2023) for my fragmentation proxy, denoted as $FRAG$, I subtract the HHI from 1 so that a higher value indicates more fragmentation. For the rest of the paper I refer to lit market fragmentation with synonymous terms such as venue concentration, venue competition or lit fragmentation.

My analysis pertains to both lit and dark fragmentation, and I therefore construct a dark fragmentation proxy to accompany my lit fragmentation proxy. Many studies empirically investigate "dark" fragmentation using alternative proxies (Hendershott and Mendelson, 2000; Hendershott and Jones, 2005; Comerton-Forde and Putniņš, 2015; Degryse, De Jong and Kervel, 2015; Foley and Putniņš, 2016; Gresse, 2017; Buti, Rindi and Werner, 2022). My measure of dark fragmentation is the same as O'Hara and Ye (2011) who use the TRF share of total volume, often denoted as a "dark-share" used in other studies. I analogously construct my dark fragmentation variable as,

$$dark_{i,t} = \frac{V_{i,t}^{D}}{\sum_{j \in K} V_{i,t,j}}, \tag{3}$$

where $dark_{i,t}$ represents the dark share for stock i on day t and $V_{i,t}^{D}$ is the total volume reported off-exchange. I refer to dark fragmentation with synonymous terms such as dark share, off-exchange share or dark fragmentation.

3.2 Liquidity Measures

I construct two variables which proxy for liquidity, the effective spread and the depth. The effective spread proxies for the cost borne to an investor who completes a round trip transaction and is computed as,

$$e_{i,t,k} = \frac{2\, q_{i,t,k}\,(p_{i,t,k} - m_{i,t,\tau})}{m_{i,t,\tau}} \times 10^4, \qquad ES_{i,t} = \frac{\sum_k V_{i,t,k}\, e_{i,t,k}}{\sum_k V_{i,t,k}}, \tag{4}$$

where e is the effective spread, q is the direction, p is the trade price and m the midpoint of trade k in stock i on day t at time τ. My trading cost proxy is denoted ES and is the volume-weighted spread. Recent literature provides compelling evidence that the securities information processor (SIP) is frequently out of order and trade signing which relies on the SIP ordering of trades and quotes will understate effective spreads (Battalio et al., 2026). I follow the remedy of Battalio et al. and construct a latency-free (LF) national best bid or offer (NBBO) by ordering all trades and quotes by their exchange time (participant time in TAQ) and sign trades against the BBO on the exchange on which the trade occurs.[21] Therefore the prevailing midpoint m and trade sign q are constructed using exchange ordered quotes as opposed to SIP ordered quotes. Given that Battalio et al. demonstrate that the SIP understatement bias is more pronounced for fragmented stocks, estimating the relation of SIP spreads and exchange fragmentation may have biased past results. I therefore estimate the relation with LF-signed spreads. The size of the data required to construct the LF-NBBO is immense; I utilize over 3 trillion exchange level quote updates in the computation.

I emphasize here that I do not include off-exchange trades in my daily value-weighted effective spread measure. Off-exchange trades face fundamentally different execution conditions and may arise from, but are not restricted to, internalized retail trades, midpoint crossing networks or dark pools. It is well documented that internalized retail trades often receive price improvement (Boehmer et al., 2021; Battalio et al., 2023; Barber et al., 2024), and that crossing networks and dark pools face lower

[21] Odd-lot trades are signed by walking back the LF NBBO by 1 millisecond and using that value of the lagged LF-NBBO midpoint. For a complete description of the signing methodology one should reference Battalio et al. (2026).

transaction costs (Menkveld, Yueshen and Zhu, 2017). If off-exchange trades are associated with lower trading costs, then my estimation of dark shares and trading costs could be mechanical, therefore I aim to only examine the transaction cost of on-exchange trades. Furthermore, the debate on market fragmentation, even when pertaining to dark fragmentation, is typically concerned with lit-exchange quality, as opposed to protecting off-exchange investor trading costs with the exception of retail traders (U.S. Securities and Exchange Commission, 2010, 2019).

I construct a second liquidity variable, total depth at the LF-NBBO. Battalio et al. (2026) do not provide instructions for how depth should be calculated using the LF-NBBO. I take a similar approach to Holden and Jacobsen (2014) and construct my depth variable denoted as $DEPTH$ by computing the time weighted size at both the bid and ask, specifically,

$$DEPTH_{i,t} = \frac{\sum_j w_{i,t,j} \left(b_{i,t,j} + a_{i,t,j}\right)}{\sum_j w_{i,t,j}}, \tag{5}$$

where b is the depth at the bid, a is the depth at the ask and w is the weight of time relative to the entire trading day. If multiple exchanges are quoting at the best price I sum the size at their respective asks and bids.

I outline my construction of control variables. I compute realized volatility by summing one-minute midpoint returns as a proxy for daily volatility. I use the mean realized volatility in cross-sectional regressions where there is one observation per stock. I construct an alternative volatility proxy, return volatility, as the rolling twenty-day standard deviation of daily close-to-close returns which I use in panel regressions following Haslag and Ringgenberg (2023). I use the median midpoint of each stock's trading day as the daily price level. Share volume is summed across stock-days and used as a proxy for activity. I compute the daily size of each stock by multiplying the shares outstanding (from TAQ master files) by the median quoted midpoint. I also create an indicator variable for whether a stock is tick constrained following Kwan, Masulis and McInish (2015).

3.3 Sample & Summary Statistics

While I aim to investigate the association of fragmentation for *most* stocks, I employ filters to retain a clean sample. First, I only study common stocks, denoted by security_type "A" in the TAQ master files, as ETFs may have fundamentally different trading characteristics (Ben-David, Franzoni and Moussawi, 2018). Second, I exclude all stocks which are "thinly-traded," using the SEC's classification as a general guideline for this restriction.[22] Through my own investigation, thinly traded stocks' HHI of trading volume can dramatically fluctuate since trading is spread thin, therefore I require that each daily observation of a stock's daily trading activity has 100,000 shares traded as well as 1,000 trades.[23] The effect of market fragmentation on thinly traded securities has been studied by Van Ness, Van Ness and Woods (2026); I therefore study stocks which are not thinly traded.[24] My third restriction is that I filter to dates between January 4, 2016 and June 18, 2026. TAQ participant timestamps were not available until late 2015 and my preferred liquidity metrics (LF-NBBO Effective spreads and LF-depth) cannot be computed before then. My final restriction is that I require at least 1,000 stock-days per symbol so the time series contains enough observations for ELW to converge.

The sample contains 1,758 stocks across 2,609 trading days for a total of 3,773,606 stock-days. Summary statistics are presented in Table 1. Despite the seemingly extensive sample restrictions imposed, I note that my sample on the average day includes 87% of total dollar volume and 89% of the market cap of all common stocks which trade. In 2022 FINRA stated the floor for a large cap stock was \$10 billion (Finra, 2022), and the median stock-day in my sample has a market cap of \$6.08 billion. The tenth percentile stock-day has a market cap of \$0.96 billion and the ninetieth percentile stock-day has a market cap of \$53.17 billion. Therefore, my sample could be characterized as the vast majority of the U.S. equities market which is not thinly traded, but not restricted to large cap stocks, with considerable size heterogeneity.

4 Exchange Venue Fragmentation Is Fractionally Integrated

4.1 Measurement and the Memory Parameter

Baillie (1996) provides a survey of the literature on long memory. A process is said to have long memory if its spectral density is proportional to λ^{-2d} as $\lambda \to 0^+$, where λ is the frequency, and the memory parameter d governs the rate at which the density diverges at the origin (Granger and Joyeux, 1980; Baillie, 1996).[25] The more intuitive characterization of long memory discussed by Baillie (1996) is a process whose autocorrelations decay at a hyperbolic rate k^{2d-1} as opposed to the geometric decay of a stationary ARMA process (ρ^k, $|\rho| < 1$).

[22] The SEC considers a stock as thinly traded if its average daily volume is less than 100,000 shares. SEC Description Click Here.

[23] Specifically, I enact the cutoff at the 1^{st} percentile of a stock's date-level time series to allow for the reduced volume on half-days or infrequent abnormally low activity trading days.

[24] Thinly traded securities are found to be particularly sensitive to market fragmentation, with quoted spreads and execution time increasing (Van Ness, Van Ness and Woods, 2026).

[25] For $d \in (0, 1/2)$ the process is covariance stationary and for $d \in [1/2, 1)$ the process is no longer covariance stationary.

Before discussing the estimation of long memory in the panel, I first provide intuition for long memory versus other time series properties. Consider three time series for General Mills (GIS) over the course of my sample: daily returns, the daily fragmentation and the daily log-price, presented in Figure 2.[26] Panel A plots the return time series of GIS, which looks like white noise [$I(0)$]; Panel B plots the fragmentation of GIS, which has visible slow moving trends but appears to mean revert [$I(d)$]; Panel C plots the log-price of GIS, where the shocks are permanent and it appears to drift with no mean reversion, evidence of a random walk [$I(1)$]. I then plot the autocorrelation function (ACF) of the cross-sectional daily returns, fragmentation and price in Panel D, Panel E and Panel F respectively. The return time series ACF decays instantly, the fragmentation time series ACF decays hyperbolically and the ACF of the random walk is near linear. We can therefore classify the venue fragmentation time series as having long memory. In Panel E and Panel F I also plot the ACF of a simulated ARFIMA$(0, \hat{d}, 0)$, and $AR(1)$ function respectively. The deviation between the simulated ARFIMA ACF and the estimated ACF is not a failure of the estimate, rather it reflects that the sample fragmentation process likely contains autoregressive and mean reversion innovations. Figure 2 only provides an example since the population ACF of an $I(\hat{d})$ process is undefined when $\hat{d} > \frac{1}{2}$, therefore the ACF in Panel E is purely descriptive and my identification is discussed below.

The autocovariance characterization of long memory is available only below the stationarity boundary. For $d \geq 1/2$ the population autocovariance function does not exist, and the characterizations that survive are the spectral one and the one in terms of partial sums. The estimates I report place venue fragmentation above that boundary, so my classification of fragmentation being a long memory process is estimated through the spectrum approach. I estimate d with the ELW estimator of Shimotsu and Phillips (2005), in the unknown-mean form of Shimotsu (2010). The estimator is semiparametric so it uses only the behavior of the periodogram at frequencies near zero, and it is agnostic to the short-run dynamics of the series. Other semiparametric estimators, the log-periodogram regression of Geweke and Porter-Hudak (1983) and the local Whittle estimator of Robinson (1995), are developed for covariance-stationary processes. I estimate venue fragmentation to be long memory with a differencing parameter $\hat{d}$ near $\frac{1}{2}$, which puts it at the edge of their validity. ELW has been used in finance literature by Chen, Eaton and Paye (2018) who adopt it for aggregate illiquidity series with $\hat{d}$ well above $\frac{1}{2}$. For a full discussion of ELW see Shimotsu and Phillips (2005).[27]

The ELW estimator requires a bandwidth parameter m where $m = n^{\psi}$, n is the number of observations in the series and $\psi \in (0, 1]$. The bandwidth parameter is the number of Fourier frequencies

[26] I select GIS as it is a recognizable name with a $\hat{d} = 0.583$, the value is near the cross-sectional median of 0.582.

[27] The `python` package `PyELW` may be helpful to the researcher looking to investigate long memory in financial time series data (Blevins, 2026).

(λ) nearest to zero that the estimator ingests. A higher m includes more frequencies and yields a more precise estimate of d, but including more Fourier frequencies away from the origin introduces more short-run dynamics.[28,29] In essence, the researcher has to balance precision while minimizing short-run dynamics. I follow Bollerslev and Jubinski (1999) and set $m = n^{0.5}$, which for $n = 2,609$ (a stock with full coverage) implies that $m = 51$, as it must be an integer.[30] This means the shortest period used to estimate d would be 51.16 days ($\frac{2{,}609}{51}$), and the longest period would be 2,609 days.[31] I often report statistics or estimate regressions at varying values of m to probe for the sensitivity of results. For the rest of the paper I use *period* and *frequency* interchangeably to refer to different Fourier frequencies; the period terminology is more analogous to extant finance terminology and is communicated in trading days. One should note that a shorter period corresponds to higher frequency and vice-versa.

Table 2 reports the estimates from ELW. I estimate the persistence of $FRAG^{lit}$ on the raw series as well as a logit transformation. The rationale is that $FRAG^{lit}$ is bounded $[0, 1)$ by construction and the long memory characterization could be spuriously identified from the series bouncing between its boundaries.[32] The logit function maps $FRAG^{lit}$ to the real number line where the boundary bouncing cannot occur. For the 1,758 stocks in the sample, the ELW estimator only converges on 1,754. The median estimate on raw (logit) $FRAG^{lit}$ is approximately 0.58 (0.59) at $m = n^{1/2}$, with an interquartile range of roughly 0.51 to 0.65 (0.52 to 0.65), and the estimate is positive for all stocks. A d above 1/2 implies non-stationarity, and below the series is still long memory but in the stationary region. My estimates imply that fragmentation is near the stationary boundary.

For a visual interpretation of ELW, in Figure 3 Panel A I plot the pooled log periodogram (the variance at different λ) of venue fragmentation and overlay the periodogram of the same series after fractional differencing by each stock's own $\hat{d}$. The variance at each frequency diverges when approaching the origin, while the differenced frequencies are relatively flat within the estimation band, which is the property ELW selects $\hat{d}$ to deliver. I tabulate a bandwidth grid in Table 2 Panel B where the median is approximately 0.60, 0.58, and 0.53 at $m = n^{0.4}$, $n^{1/2}$, and $n^{0.6}$. I also present $\hat{d}$ with the IQR across varying bandwidths in Figure 3 Panel B, as the bandwidth increases $\hat{d}$ decreases and

[28]The estimator may also fail to converge with a low m as there are too few frequencies.

[29]Lobato and Velasco (2000, p. 418) note "*it is obvious that inference is more accurate as m is larger. For big values of m, however, the statistics employed would reflect the medium- and short-term behavior of the process that would bias our inference. This trade-off between bias and variance was addressed, for instance, by Robinson (1994b) and Henry and Robinson (1996) for the univariate case.*" and present their results for a grid of m.

[30]Bollerslev and Jubinski (1999, p. 18) call $n^{1/2}$ "our reported value" of the truncation parameter and vary it to $0.5n^{1/2}$ and $1.5n^{1/2}$ as a sensitivity check, finding that "although the quantitative results vary slightly from case to case, the qualitative results do not." They estimate d with the log-periodogram regression of Geweke and Porter-Hudak (1983) rather than with ELW.

[31]Stocks which have fewer observations in their time series would still use 51 frequencies, but the period of each frequency would be shorter depending on the number of observations.

[32]Estimating fractional integration on bounded share-type series has been done (Byers, Davidson and Peel, 1997).

the IQR tightens. The magnitude of memory varies with the bandwidth grid, while the finding that venue fragmentation carries long memory is consistent. I also present the cross-sectional distributions in Figure 3 Panel C.

Spurious long memory can arise from a number of sources (Velasco, 1999; Diebold and Inoue, 2001; Granger and Hyung, 2004; Shimotsu, 2010; Qu, 2011). I extensively test the data for true versus spurious long memory in Appendix A. I find that while the stock-level time series do possess some structural breaks, which is not unexpected given the dynamic nature of equities markets, there is indeed true long memory in the stock-level series across the sample. The estimations are consistent with simulated long memory.

I next investigate whether there is persistence identified with other stock-level time series in my data. In Panel C of Table 2 I report $\hat{d}$ for volume, ES, $DEPTH$ and $dark$. I find $\hat{d}^{volume} = 0.50$, $\hat{d}^{ES} = 0.66$ and $\hat{d}^{DEPTH} = 0.75$, indicating all are long memory processes on the stock level, consistent with Lobato and Velasco (2000) for volume and loosely consistent with Chen, Eaton and Paye (2018) who find aggregate illiquidity is long memory. Not only are dimensions of liquidity long memory, but I estimate ($\hat{d}^{dark} = 0.57$) is a long memory process on the stock level. The finding of persistence in off-exchange share is related to Kwan, Masulis and McInish (2015) who state that order flow migrates to dark venues when the probability of trade execution rises there, creating a positive feedback loop.

4.2 Aggregation and the Common Component

Motivated by the findings of Chen, Eaton and Paye (2018) that aggregate illiquidity is long memory, I now examine whether aggregate fragmentation is long memory. I also examine whether long memory of aggregate liquidity variables is identified in my data as it was in Chen, Eaton, and Paye's sample. I plot the equal-weighted average of $FRAG^{lit}$, $dark$, ES and $DEPTH$ across stocks in Figure 4 for reference. The cyclical drops and spikes are predominantly triple witching Fridays.[33] A slow drift with mean reversion is visible in all the panels. I tabulate the ELW estimates on the aggregate series in Table 3 using $m = n^{1/2}$. The differencing parameters in the aggregate series are $\hat{d}^{FRAG} = 0.73$, $\hat{d}^{dark} = 0.74$, $\hat{d}^{ES} = 0.61$ and $\hat{d}^{DEPTH} = 0.74$. The two aggregate fragmentation time series appear more persistent than the stock-level series, while the aggregate liquidity series and the stock-level series appear similar. I also make note that dark fragmentation has risen to near 50% of the equal weighted volume in my sample.

The identification of a seemingly more persistent common time series raises the question of whether the stock-level series inherit their documented persistence from the aggregate component. I

[33] All aggregate and stock-level memory estimates are robust to back-filling or forward-filling triple witching days.

follow Ergemen and Velasco (2017) who extend the common factor projection of Pesaran (2006) to non-stationary factors.[34] I then reestimate the ELW on the projected series which is orthogonal to any common factors by design. The median $\hat{d}$ of the projected series on the stock level is presented in Table 3 using a bandwidth of $m = n^{1/2}$. Without performing a test for statistical differences, the projected series of $FRAG^{lit}$and $dark$ appear to be less persistent than the raw series estimates from Table 2, while the two liquidity series appear similar. I then compute the difference between the aggregate series estimate of $\hat{d}$ and the stock-level median estimate of $\hat{d}$ and standardize the difference using the standard error from the aggregate series. The comparison is descriptive in nature, and is suggestive that the projected series of the fragmentation variables appear less persistent than the aggregate fragmentation series, while the liquidity series appear to possess a similar persistence to aggregate counterparts. While a common component in liquidity is not a novel finding (Chordia, Roll and Subrahmanyam, 2000; Hasbrouck and Seppi, 2001), nor is its persistence (Chen, Eaton and Paye, 2018), I find that fragmentation also contains a common persistent component.

Table 3 suggests that the stock-level fragmentation series may inherit a part of their persistence from the common component. I do not perform a statistical test to confirm the inheritance of memory on the stock-level series from the aggregate series since Pesaran (2015, 2021) note the difficulties in accurately identifying cross-sectional dependence in fractionally integrated time series. However, I treat inheritance of a common factor into the stock-level time series as a possibility that must be recognized in any estimation of the relation between fragmentation and liquidity.

4.3 Fractional Cointegration between Fragmentation and Liquidity

I carry out a diagnostic outlined by Engle and Granger (1987) to determine whether fragmentation and liquidity are cointegrated. The finding for or against cointegration would dictate subsequent econometric methods to analyze the relation between the two variables. I regress each liquidity measure on each fragmentation measure and a constant in levels. Specifically I regress $y_t = a + bx_t + \varepsilon_t$ where $y \in \{ES, DEPTH, dark, FRAG^{lit}\}$ and $x \in \{FRAG^{lit}, dark, volume\}$, I then estimate the memory of the residuals ε_t by ELW at $m = n^{1/2}$. A requirement for cointegration is that the residual must be less persistent than both series, so the term, denoted the "reduction," $\min(\hat{d}^x, \hat{d}^y) - \hat{d}^\varepsilon$ being positive would imply cointegration and being negative would not. Another empirical curiousity is wheether lit and dark fragmentation's persistence are cointegrated, so I include the two as a pair. I also inspect the reduction of both fragmentation measures with volume. Since both lit and dark fragmentation are constructed using volume shares there is a concern that the persistence

[34] For a full explanation of Ergemen and Velasco (2017) see Section 6.

is just a repackaging of the already accepted persistence of volume (Bollerslev and Jubinski, 1999; Lobato and Velasco, 2000).

The regression and ELW estimation on residuals is done stock-by-stock and reported in Panel A of Table 4. The median reduction is negative for all four fragmentation-liquidity series pairs, ranging from -0.077 to -0.145, and the residual is more persistent than the less persistent of the two series for between 71.7% and 88.5% of stocks. The fragmentation-volume pairs also present residuals which are more persistent than either of the series. The evidence is in stark disagreement with the outline for cointegration. Furthermore, Engle and Granger (1987) note this test is biased towards finding evidence for cointegration.

In Panel B of Table 4 I repeat the diagnostic on the aggregate series. Since I do not have a cross-section to compute descriptives, I perform a sensitivity analysis of the bandwidth parameter m selected. The bandwidth sensitivity is applied to both the ELW estimation identifying the aggregate series $\hat{d}$ but also the $\hat{d}^{\varepsilon}$. I report the reduction along with the reduction in standard errors where the standard error is equal to $\frac{1}{2\sqrt{m}}$ (Robinson, 1995). The largest positive reduction is $+1.42$ standard errors, for lit fragmentation and the effective spread at $m = n^{0.4}$. At $m = n^{1/2}$ no pair reaches a single standard error in either direction. At $m = n^{0.6}$ the two depth pairs are significantly negative, -3.07 and -3.63. I therefore find no evidence of fractional cointegration between any fragmentation liquidity pair, or fragmentation volume pair at the stock level or in the aggregate series.

5 The Cross-Section of Persistence

My estimates from Section 4 assert that stocks have persistent fragmentation time series. In addition $\hat{d}$ appears normally distributed across stocks. I investigate if stock characteristics are related to the distribution of $\hat{d}$. For brevity, the exhibits in the section will have $FRAG^{lit}$as the object of interest and *dark* exhibits are presented in the online appendix with a brief discussion at the conclusion of this section. I note that the estimates in this section are purely cross-sectional correlations. First, I estimate

$$\hat{d}_i^{FRAG} = \alpha + \boldsymbol{\beta}'\mathbf{Z}_i + \varepsilon_i, \tag{6}$$

using one observation per stock, where $\hat{d}_i^{FRAG}$ is the ELW estimate for stock i from Section 4 and $\mathbf{Z}_i$ is a vector of seven stock characteristics, each measured as a stock mean over the sample window. The characteristics are: *inverse price*, the inverse of price, in log units; *volatility*, computed as the sum of squared minute-level midpoint returns, in log units; *size*, measured as the shares outstanding multiplied by the median midpoint of that day, in log units; *volume*, the daily share volume, in log units; *dark*, the share of volume off-exchange; *tick constrained*, an indicator for tick constrained stocks, constructed following the definition of Kwan, Masulis and McInish (2015); and *fragmentation*, the raw fragmentation level.[35] I report coefficients with heteroskedasticity-robust standard errors throughout.

Table 5 reports the estimates. I first enter each variable on its own in a univariate setting in columns (1)–(7). Taken one at a time, the table reveals that higher persistence in $FRAG^{lit}$is associated with lower priced, more volatile, smaller, higher volume, tick constrained and more fragmented stocks. Once entered jointly in column (8), the table suggests that the characteristics may be capturing similar information. Only size and volume retain their significance, while the dark share becomes significant and is estimated with a negative sign. Consistent with variables capturing similar information, I present the VIFs from the specification estimated in column (8), in column (11). Given that a VIF greater than 5 or 10 typically implies collinearity, and market cap is associated with the greatest VIF recorded at 12.2, I drop market cap from the regression and estimate the reduced specification in column (9). The VIFs from the reduced specification are presented in column (12) and all are below 3. The reduced specification in (9) reveals that lower priced, higher volume and lower dark share stocks are associated with higher persistence in $FRAG^{lit}$.

I discuss in subsection 4.1 that there is a trade-off regarding the bandwidth parameter, where a lower bandwidth introduces relatively noisier estimates but incorporates more long period variance relative to short period variance. A variable could be said to have a "pure" association with the low-frequency activity of $FRAG^{lit}$if it is estimated with a significant sign using a smaller bandwidth, where the estimate is least contaminated by short period activity. I reestimate the regression using a bandwidth of $m = n^{0.4}$ and tabulate the coefficients in column (10). *Inverse price*, *volume* and *dark* retain their estimated signs and significance while *fragmentation* becomes significant and positive, a reversal of the negative sign it carries in column (8).[36] To better understand the relation of the variables and $\hat{d}^{FRAG}$ I plot the coefficients and standard errors across the range of bandwidths in Figure 5. I find that while *inverse price*, *volume* and *dark* appear to possess some high-frequency dynamics, the relation is significant at the lowest bandwidths and does not appear to be contaminated by the high-frequency dynamics. I find that *fragmentation* appears to have two competing forces on either end of the band spectrum. At the lower-frequencies the raw level of fragmentation is positively related to the persistence of $FRAG^{lit}$, but the higher-frequencies are dominated by a negative relation.

[35] *Inverse Price, volume* and *size* are the stock-level mean of daily log observations, whereas *volatility* is the log of the stock-level mean. The difference is because 0.02% of daily observations have zero *volatility.* Coefficients are similar for *Inverse Price, volume* and *size* when computed as the mean of daily logs.

[36] I choose $m = n^{0.4}$ as the lowest end of the range as below this for many stocks the ELW estimator fails to converge.

This implies that the persistence relation estimated from the wider bandwidths is likely contaminated by high-frequency dynamics. To summarize, the memory of $FRAG^{lit}$is associated with lower priced, higher volume stocks, lower dark share and more fragmented stocks.

I estimate the same regression and bandwidth sensitivity where $dark$ is the dependent variable; the exhibits are presented in Table A3 and Figure A5 respectively. The persistence in dark share appears to be associated with higher priced, higher volume, higher dark share, tick constrained and higher lit fragmentation. Kwan, Masulis and McInish (2015) find that traders may go off-exchange for execution to bypass the tick constraint in high-priced stocks and this may create a self-reinforcing spiral as the probability of off-exchange execution increases as more traders migrate. In line with their findings, I find that the persistence in dark share is related to high-priced and tick constrained stocks. I also investigate in the online appendix whether the persistence in the varying fragmentation and liquidity measures is associated with one another in the cross-section while controlling for stock characteristics (Figure A6, Table A4). I find the persistence of $FRAG^{lit}$is positively associated with the persistence of volume, ES and $DEPTH$. These findings are intuitive: fragmentation is constructed from volume shares, and a trader's routing choice is endogenous to displayed liquidity. I do not find that the persistence of $FRAG^{lit}$and $dark$ nor $dark$ and volume are related.

I make a key point using the estimations of this section along with the finding that both facets of market fragmentation appear to possess both stock-level and aggregate components in subsection 4.2. Given that stock characteristics are related to the level of persistence, the notion of heterogeneous loadings on the stock level of common factors which may drive the persistence in fragmentation remains a possibility. This notion influences the empirical methods selected to estimate a relation of fragmentation and liquidity in the panel.

6 Estimation and Inference under Long Memory

In subsection 3.1 I note that I construct my lit fragmentation measure $FRAG^{lit}$using the volume shares of all lit exchanges and exclude exchange "D" which aggregates all off-exchange volume reported to a TRF. My choice of variable differs from recent work (Haslag and Ringgenberg, 2023), where the fragmentation measure includes the volume share of off-exchange activity. I first motivate my construction of $FRAG^{lit}$before I proceed to my desired econometric specification. I estimate regressions in the following form,

$$Y_{i,\tau} = \beta_1 FRAG^{T}_{i,t} + \beta_2 dark_{i,t} + \gamma Z_{i,t} + \alpha_i + \nu_t + e_{i,t}, \tag{7}$$

where $Y_{i,\tau} \in \{ES_{i,\tau}, DEPTH_{i,\tau}\}$, $\tau \in \{t, t+1\}$ and T represents the two types of stock-day fragmentation measures constructed using all volume shares, or only lit volume shares. $Z_{i,t}$ is a vector of controls which includes: the daily share volume, in log units; the daily inverse price, in log units; the 20-day rolling standard deviation of daily log returns, in log units; the daily market capitalization, in log units. Outcome variables, fragmentation variables and controls are all winsorized at 0.5/99.5 percentiles. I also include stock and date fixed effects denoted by α_i and ν_t respectively. The specification and controls are similar by design to Haslag and Ringgenberg (2023). Directly comparing to their specification I do not include market-to-book and leverage as controls; I add contemporaneous volume as a control; and I estimate outcomes both contemporaneously and in the next period whereas they only estimate outcomes one period forward. I first enter $FRAG^T$ with controls; I then add *dark*, repeating for both fragmentation measures, periods and outcome variables. I estimate the regressions on t and $t+1$ outcomes as contemporaneous outcomes are my preferred design choice which I will use subsequently, but $t+1$ is used in Haslag and Ringgenberg (2023).

The results are presented in Table 6. Column (1) estimates a negative contemporaneous relation between *ES* and $FRAG^{all}$, but when including *dark* in column (2) the coefficient is insignificant with *dark* estimated with a positive sign. In column (3) I estimate a positive relation between *ES* and $FRAG^{lit}$. When including *dark* in column (4) the coefficient on $FRAG^{lit}$ slightly decreases but retains sign and significance and the coefficient on *dark* is similar in sign and significance to (3). I then re-estimate the same models in columns (5)-(8), except the outcome is now $ES_{i,t+1}$. I find a similar pattern: the coefficient on $FRAG^{all}$ flips in sign from (5) to (6) and is statistically significant in both, whereas $FRAG^{lit}$ remains stable when *dark* is included.

In panel B of Table 6 I repeat the same process with *DEPTH* as the dependent variable. The coefficient on $FRAG^{all}$ does not flip signs or lose significance when including *dark*, but *dark* is estimated with a positive relation when entered with $FRAG^{all}$ as opposed to a negative relation when entered with $FRAG^{lit}$. I note here that the $\rho(FRAG^{all}, dark) = -0.8269$ whereas $\rho(FRAG^{lit}, dark) = -0.0066$ and finally $\rho(FRAG^{lit}, FRAG^{all}) = 0.4438$. Constructing the HHI of market fragmentation using all exchanges as Haslag and Ringgenberg (2023) do creates a fragmentation measure that appears to include relatively more information about the off-exchange share than it does about lit fragmentation. I am explicit here that the purpose of this demonstration is not adversarial. The object of Haslag and Ringgenberg's study is overall market fragmentation, therefore their measure is correct as used. But as I have just shown, using all exchange volume shares bundles two effects which may have fundamentally different relations to liquidity. I wish to estimate the relation of lit and dark fragmentation separately and for the rest of the paper I use $FRAG^{lit}$ to proxy for lit fragmentation and *dark* for dark fragmentation.

I conclude my discussion of Table 6 by summarizing the relation of market fragmentation and liquidity, under a standard econometric specification (two-way fixed effects with two-way clustered standard errors). Lit fragmentation is associated with wider spreads, but increased depth. Dark fragmentation is associated with wider spreads and decreased depth. These associations are documented whether the outcome is liquidity on the current period or one period forward. I reaffirm here that both liquidity measures are calculated using only on-exchange spreads and depth, therefore a change in the dark share cannot mechanically alter trading costs, as off-exchange trades may have fundamentally different execution costs (Menkveld, Yueshen and Zhu, 2017). I refer to the relations from Table 6 as the *levels* relation between market fragmentation and liquidity and return to them when synthesizing these statistics with later findings.

I now move forward to outline the settings of my study and discuss how these settings influence the econometric specification I employ. Recall that during the sample aggregate and stock-level fragmentation in addition to aggregate and stock-level liquidity are identified as long-memory processes. If fragmentation and liquidity load heterogeneously on correlated persistent common factors, date fixed effects leave residual factor components in both variables and the estimator can be inconsistent (Pesaran, 2006; Bai, 2009). While the persistence in the common factor is not the root of the issue, persistence can perpetuate the problem. This is because in a persistent time series variation concentrates at long periods, and under heterogeneous loadings the surviving factor component dissipates slowly. This accumulation of unabsorbed variance enters the regression as an omitted variable and it may be influential in the sample covariance observed to estimate the coefficients (Ergemen and Velasco, 2017). To reiterate, I do not prove heterogeneous loadings exist in the data. What I do prove is that a common persistent factor exists in the data and I choose an estimator that is robust whether the stock-level loadings on said factor are heterogeneous or homogeneous.

I employ the methodology of Ergemen and Velasco (2017) (henceforth EV), who extend Pesaran (2006) to panels in which the outcome, regressors, error terms, and the common factors are all fractionally integrated. EV establish consistency and asymptotic normality of the slope estimates in this setting. Following EV I begin by taking first differences, which removes the stock fixed effects and reduces the memory of every series.[37] I then include the cross-sectional averages of the differenced outcome, fragmentation measure, dark share, and controls as additional regressors, each with a stock-specific

[37]Their method permits additional fractional differencing beyond the first difference, using the parameter $\delta^* \geq 1$, utilized when the data are persistent enough to require it. I set $\delta^* = 1$, their own choice in both their simulations and their application, and the smallest value their conditions permit given the memory I estimate in Section 4.

coefficient.[38] Specifically, I estimate

$$\Delta Y_{i,t} = \alpha_i + \beta_1 \Delta FRAG^{lit}_{i,t} + \beta_2 \Delta dark_{i,t} + \gamma' \Delta Z_{i,t} + \phi_i' \bar{H}_t + \varepsilon_{i,t}, \tag{8}$$

where Δ indicates a first difference, $\bar{H}_t$ contains the day t cross-sectional averages of ΔY_t, $\Delta FRAG^{lit}_t$, $\Delta dark_t$ and ΔZ_t. The averages are the proxy for the unobserved factor, and ϕ_i is estimated on the stock-level, which yields the heterogeneous loading. The control variables in ΔZ_t include: the daily share volume, denoted *volume*, in log units; the inverse price, denoted *price* in log units; the daily volatility, computed as the rolling 20-day standard deviation of log returns, denoted *volatility*, in log units; and the market capitalization, denoted *size* in log units. I report the pooled estimate, in which β_1, β_2 and γ are common across the 1,758 stocks in the panel and standard errors clustered by stock and date.[39]

Before reporting the results, I concede that while EV corrects for persistence and the common factor, it only recovers an association and does not deliver causality. If liquidity determines trading activity and where orders are sent, the two may be jointly determined. To address this I estimate the relation using the identification strategy of Degryse, De Jong and Kervel (2015) (henceforth DDK). They instrument a stock's fragmentation and dark share with the average fragmentation and dark share across the other stocks in the same size group on the same day. The instrument is relevant because institutional investors trade several stocks across alternative venues on the same day, so fragmentation is correlated across stocks, and it is excludable because it is unlikely the liquidity of one stock influences the fragmentation of other stocks in the same group. I estimate

$$FRAG^{lit}_{i,t} = a_{1,i} \times d_{q(t)} + b_1' X_{i,t} + c_1' W_{i,t} + \epsilon_{1,i,t}, \tag{9}$$

$$(FRAG^{lit}_{i,t})^2 = a_{2,i} \times d_{q(t)} + b_2' X_{i,t} + c_2' W_{i,t} + \epsilon_{2,i,t}, \tag{10}$$

$$dark_{i,t} = a_{3,i} \times d_{q(t)} + b_3' X_{i,t} + c_3' W_{i,t} + \epsilon_{3,i,t}, \tag{11}$$

$$Y_{i,t} = \alpha_i \times \delta_{q(t)} + \beta_1 \widehat{FRAG^{lit}}_{i,t} + \beta_2 \widehat{(FRAG^{lit})^2}_{i,t} + \beta_3 \widehat{dark}_{i,t} + \gamma' W_{i,t} + e_{i,t}, \tag{12}$$

where $X_{i,t}$ are the three leave-one-out size group averages, $FRAG^{lit}$, $(FRAG^{lit})^2$ and $dark$.[40] There are four size groups formed based on quartiles of the mean size over the whole sample; stocks do not

[38] This is the step which projects out the common factor and is effectively date fixed effects with heterogeneous loadings.

[39] EV define the pooled estimator but leave its asymptotic analysis to future research; they report the mean group (MG) estimates and provide the asymptotic theory for MG only. I choose to report the pooled estimate because it is analogous to the pooled specifications I compare it against. The MG estimates for the same model are equivalent in sign and statistical significance across both *ES* and *DEPTH*. The coefficient in magnitude is slightly larger for *DEPTH* and slightly smaller for *ES*.

[40] $(FRAG^{lit})^2$ is the average of the squares not the square of the average.

change groups. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as $AvgLiq$; market cap, in log units, denoted as $Size$; the inverse price, in log units, denoted as *Inverse Price*; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable *Algo* in DDK's specification is standardized message traffic. As I do not have message data, I proxy with NBBO updates divided by trading volume, but note the limitation. The specification also entails $\alpha_i \times \delta_{q(t)}$ which are stock-quarter fixed effects. I estimate the specification by two-stage GMM as they do, with standard errors clustered by stock rather than the Newey-West errors they report. The clustered standard errors account for the serial correlation in the error terms that is present given that the time series of the regressors and outcomes are long memory. I report the marginal effect at the sample mean as well as the 10^{th} and 90^{th} percentiles. Another deviation I make related to their methodology is to fractionally difference every series in the system which includes the outcomes, the three endogenous variables, the three instruments and the controls; this is done to account for the long memory property I document.[41] I note here that the marginal effects at the mean, p10 and p90 with and without the differencing are identical in sign and significance, reported in the online appendix Table A6, Table A7. Utilizing the DDK specification is not meant to be a replication of their study to scrutinize their coefficients. The sample, time period and markets (Europe vs. U.S.) used in this paper and DDK differ. The comparison is EV and DDK which is meant to illustrate the difference between a design that strips a common factor (EV) and a design that effectively instruments on the common factor (DDK). I demonstrate later in this section that the relation of fragmentation and liquidity is conditional on two dimensions, the period and component (idiosyncratic versus common). Both of these specifications implicitly select on each of these dimensions and therefore are both *correct*, but only in explaining one section of the entire relation.

The results from both specifications are presented in Table 7. Columns (1)-(2) present the coefficients for the EV estimations, columns (3)-(4) present the coefficients for DDK without the quadratic term, and columns (5)-(6) present the coefficients estimated with the quadratic term. In columns (5)-(6) I also present the marginal effects at the mean, p10 and p90. Control variable coefficients are suppressed for brevity; the online appendix presents full tables for: DDK with the quadratic term in levels and fractionally differenced Table A6, DDK without the quadratic term in levels and fractionally differenced Table A7, and EV Table A8. The EV estimation in (1)-(2) reveals

[41] Each stock's series are filtered at that stock's own $\hat{d}^{FRAG}$. The filter is applied before the stock-quarter demeaning. When fractionally differencing I remove the first 250 observations of each stock's series because the filter is one-sided and truncated where the series begins. DDK specifications are performed without the filtering and retain all observations in Table A6, Table A7. Results are insensitive to filtering.

positive associations between lit fragmentation and liquidity, whereby spreads are tightened and depth is increased. The DDK estimation of lit fragmentation without the quadratic term in (3)-(4) yields opposite signed slope estimates to EV. The estimate for ES with the quadratic term (column 5) has an insignificant marginal effect at the mean as the mean is near the global max of the curve, whereas $DEPTH$ is estimated with a negative marginal effect throughout the curve.[42] Turning to the coefficient on *dark*, across the specifications *dark* is estimated with opposite signed coefficients, all statistically significant. The two designs therefore disagree on all four relations; I next investigate whether the root of the disagreement is related to the controls, memory correction or the instruments.

I perform a series of tests to identify the disagreement. Adding *AvgLiq* and *Algo* to the EV specification does not alter the coefficients (unreported). Estimating DDK on levels as opposed to fractionally differencing retains similar estimates for the linear specification Table A7. Estimating DDK on levels for the quadratic specification does not change the marginal effects at the mean, p10 and p90 Table A6. Therefore it does not appear to be the controls or memory correction that drive the difference between EV and DDK. I then hold the FE, differencing and controls constant in DDK and alter the instruments in Table A9. I find this last inspection to reveal what causes the disagreement. Averaging over the size group as an instrument removes stock-specific variation and leaves what the group shares, the *common* component. This means instrumenting does not recover an exogenous (relative to liquidity) measure of the stock's own routing, it replaces the idiosyncratic component with the common one. The OLS and 2SLS specifications from Table A9 therefore identify different objects rather than the same object with different biases. The coefficient on *dark* flips sign when instrumented with *dark* alone, implying that the idiosyncratic and common relation of dark with liquidity may be opposite in sign, which I confirm following this. $FRAG^{lit}$does not flip when instrumented alone, implying that the common and idiosyncratic components of $FRAG^{lit}$should agree in sign, which I also confirm following this. $FRAG^{lit}$exhibits the sign flip only when instrumented with both $FRAG^{lit}$and *dark*, not either alone. While not tabulated, I find $FRAG^{lit}$ and *dark* correlate +0.134 as regressors (idiosyncratic) but the leave-one-out size group averages (common) correlate +0.631. The first stage therefore replaces two nearly orthogonal regressors with two near-copies of one *common* factor, and the second stage is then tasked to split that one series across two coefficients, which it cannot separate.[43]

The coefficients from the EV specification are attractive. Often a researcher desires stationarity in the data and to strip out common effects. The EV specification performs both these functions on the

[42] I display the curve and marginal effects in Figure A3.

[43] The sum of the coefficients on $FRAG^{lit}$and *dark* when regressors in the 2SLS specification for either dependent variable remain similar (ES -0.316 $\rightarrow$ -0.390, or $DEPTH$ 0.203 $\rightarrow$ 0.252), while each individual coefficient reverses. The combined effect appears stable while the attribution of the effect between the regressors shifts.

panel and presents coefficients that describe the idiosyncratic relation between market fragmentation and liquidity, although I concede it does not identify causality or direction. If one were to seek out a coefficient which measures the relation of fragmentation and liquidity without any contamination, EV would reflect this. If the analysis were to end here, I would report that lit fragmentation improves liquidity whereas dark fragmentation deteriorates liquidity. I do not conclude my analysis with the EV estimation as the earlier findings motivate a more intricate investigation into the relation between fragmentation and liquidity.

Recall that both dimensions of fragmentation (lit and dark) and liquidity (spreads and depth) are long memory processes with a persistent common component. A property of a long memory process is that variance is not distributed equally across the frequency domain. I quote Engle (1974, p. 5) directly: *"... it may be too much to ask of a model that it explain both slow and rapid shifts in the variables ... it is at least reasonable to test the hypothesis that the same model applies at various frequencies ... since the typical spectral shape of economic variables has a strong peak at low frequencies, these periodogram components completely dominate the parameter estimates."* To fit different coefficients to the different frequencies within the panel, I use a fast Fourier transform (FFT) to move all series from the time dimension to the frequency spectrum. Müller and Watson (2015) discuss how a fixed-length time series only holds a few low-frequency projections and analyzing long run dynamics is a small-sample problem. By pooling across stocks, I circumvent the small-sample issue. Similar spectral regressions have been used recently in asset pricing literature to determine whether factors vary in the spectral domain; I do not directly follow their methods as they assume stationarity of the data (Bandi et al., 2021) which my time series violate. My final empirical examination involves investigating the relation of fragmentation and liquidity across the frequency (I use frequency and period interchangeably) domain, and to split the relation between the idiosyncratic and common components.

I first loop through all stocks' time series data with an FFT, which yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \dots, \lfloor\frac{n}{2}\rfloor\}$ and n is the number of daily observations for stock i.[44] I then pool the periods into ten different bands, ranging from periods of 2-3 days to periods of 500+ days.[45] I next set the Fourier coefficients outside a pooled band to zero and invert the transform, which returns each stock's series filtered to that band. Each panel is the same number of stock-day observations, with fitted values based on the waves in the band. I then estimate OLS on each fitted panel for the bands

[44]I impose a further sample restriction at this point; I only retain stocks which have greater than 500 daily observations. This reduces the sample from 1,758 stocks to 1,672 stocks. The purpose of this restriction is that stocks will contribute similarly in their periods to the pooled bands.

[45]The full ranges of the ten bands are [2,3), [3,5), [5,10), [10,20), [20,40), [40,62), [62,125), [125,250), [250,500), [500, 2,609) days.

following Engle (1974). In Appendix B I present diagrams which display the conversion of time series data to the spectral dimension for contextualization of my methodology in this section.. Specifically, within each band I estimate

$$Y_{i,t}^{\mathcal{B}} = \beta_1^{\mathcal{B}} FRAG_{i,t}^{lit,\mathcal{B}} + \beta_2^{\mathcal{B}} dark_{i,t}^{\mathcal{B}} + \gamma^{\mathcal{B}\prime} Z_{i,t}^{\mathcal{B}} + \varepsilon_{i,t}^{\mathcal{B}}, \tag{13}$$

where $\mathcal{B} \in \{1, 2, \ldots, 10\}$, representing the fitted observations within a band $\mathcal{B}$; $Z^{\mathcal{B}}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. Standard errors are clustered on stock and on date. Each filtered stock series is mean zero because the zero frequency is discarded, which takes the place of stock fixed effects. Date fixed effects are omitted by design as I decompose the band coefficients into a common and idiosyncratic component afterward.

I present the coefficients of $FRAG_{i,t}^{lit,\mathcal{B}}$ and $dark_{i,t}^{\mathcal{B}}$ across bands in Figure 6 with a dotted line representing the pooled coefficient. The coefficients are tabulated in Table A10. $FRAG^{lit}$ is associated with tightened spreads for all periods under 125 days, but with widened spreads for all periods greater than or equal to 125 days. Turning to depth, $FRAG^{lit}$ is associated with a deeper book at all periods. The association of *dark* and *ES* is near zero from periods equal to or under 40 days, although the relation of off-exchange activity and spreads is positive in the longer period bands, implying wider spreads. Interestingly, the relation of *dark* and *DEPTH* is negative at periods of 250 days or less, but flips at periods greater than 500 days. I then perform a statistical test to ensure that the band coefficients are indeed statistically different from one another. Engle (1974) states the null that a coefficient is common to every band. I test the null in Table A12 under the two-way clustered covariance matrix. The null is rejected at the 1% level for $FRAG^{lit}$ and *dark* separately and jointly, for both *ES* and *DEPTH*. Restricting all coefficients to be equal across the ten bands yields a clustered χ^2 of 8,283.4 for *ES* and 3,347.4 for *DEPTH*, on 54 restrictions. The test rejects the null and implies that there are statistically significant differences across the bands. As stated, there are no time fixed effects, so both common effects and idiosyncratic effects contribute to the estimated relations across the bands. I now partition the common effects and idiosyncratic effects and estimate separate coefficients.

The presence of a common component in liquidity is well established. Chordia, Roll and Subrahmanyam (2000) document that individual stocks' spreads and depths co-move with market-wide liquidity, and Hasbrouck and Seppi (2001) decompose Dow stocks' returns, order flows and liquidity into common and stock-specific parts. As with the EV specification, I avoid forcing homogeneous loadings on a common factor and estimate a heterogeneous loading ϕ_i^z stock by stock. Let $z_{i,t}$ denote any series in Equation 13, so that $z \in \{Y,\ FRAG^{lit},\ dark,\ Z\}$, and let $\ddot{z}_{i,t} = z_{i,t} - \bar{z}_i$ be that series after removing its stock mean $\bar{z}_i$. Following Pesaran (2006) I proxy the common component with the day-t cross-sectional averages of every variable,

$$\bar{h}_t = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} (\ddot{Y}_{i,t},\ FR\ddot{A}G_{i,t}^{lit},\ da\ddot{r}k_{i,t},\ \ddot{Z}_{i,t}{}')', \tag{14}$$

where $\mathcal{N}$ is the set of stocks observed on each day t and N_t is its size. Then regress each series at the stock level,

$$\ddot{z}_{i,t} = a_i^z + \phi_i^{z\prime}\bar{h}_t + u_{i,t}^z, \qquad \underbrace{z_{i,t}^c = \hat{a}_i^z + \hat{\phi}_i^{z\prime}\bar{h}_t,}_{\text{common}} \qquad \underbrace{z_{i,t}^\iota = \hat{u}_{i,t}^z}_{\text{idiosyncratic}}. \tag{15}$$

The common and idiosyncratic components are orthogonal by design. I then apply the FFT to the common filtered panel and idiosyncratic filtered panel separately. Lastly, I pool across the same 10 bands from Equation 13 and estimate,

$$Y_{i,t}^{\mathcal{B},s} = \beta_1^{\mathcal{B},s} FRAG_{i,t}^{lit,\mathcal{B},s} + \beta_2^{\mathcal{B},s} dark_{i,t}^{\mathcal{B},s} + \gamma^{\mathcal{B},s\prime} Z_{i,t}^{\mathcal{B},s} + \varepsilon_{i,t}^s, \qquad s \in \{c, \iota\}, \tag{16}$$

where s denotes the common (c) or idiosyncratic (ι) component of the FFT filtered data pooled over band $\mathcal{B}$. Each regression is run separately on the outcome-band-component level, similar to Bandi et al. (2021). Since the components are orthogonal by design, the coefficients are similar to what would be estimated by running the regression with the outcome variable as the sum of both the common and idiosyncratic components $Y_{i,t}^{B,c+\iota}$ while each regressor variable is entered as the common and idiosyncratic component separately.[46]

The coefficients are presented in Figure 7 with each panel displaying the coefficients for the common and idiosyncratic components along with a dotted horizontal line for the pooled coefficient. I state the estimated associations here and reconcile the results with existing literature in Section 7. The relation between *ES* and $FRAG^{lit}$ is negative for both the idiosyncratic and common components at periods under 125 days, whereas it is positive for both components at periods longer than 125 days. The magnitude of the coefficient for $FRAG^{lit}$ when *ES* is the dependent variable is generally larger for the common component than the idiosyncratic component. The relation between *DEPTH* and $FRAG^{lit}$ is positive for both components, for all periods, and the magnitudes are similar across

[46] I present a figure estimating the regression in this manner in Figure A4. One might expect the coefficients to be identical given the orthogonal construction on the stock level. The estimates are not identical as the FFT filter can violate the orthogonality; other papers avoid this, but their methodology requires stationarity, which I do not have in my data (Bandi et al., 2021; Bandi and Tamoni, 2023). Regardless, results do not appear affected.

components. The association of *dark* and *ES* is positive in all periods for the idiosyncratic component whereas for the common component it is negative for periods 40 days or less and positive for periods of greater than 40 days. Lastly, the coefficient measuring the association between *dark* and *DEPTH* is positive for the common component at all periods and negative for the idiosyncratic component at all periods. The relation between fragmentation (lit or dark) and liquidity (*ES* or *DEPTH*) differs across the frequency spectrum and across components. This yields a relation which is dependent on which sector of a frequency $\times$ component spectrum is observed.

7 Discussion

I now compare the coefficients from the levels, DDK and EV specifications to what is estimated across the band spectrum. The levels specification in Table 6 column (4) effectively estimates the pooled relation indicated by the dotted line in Figure 7. The difference in magnitudes of the coefficients is a product of not including date fixed effects in the regressions estimated in Figure 7. In unreported results the coefficients are near exact when using date FE. The pooled/levels estimation of the relation between fragmentation and liquidity conceals significant heterogeneity across the frequency $\times$ component spectrum. It is unclear what a single coefficient expresses, if anything, about the relation of fragmentation and liquidity. I next pinpoint the findings of DDK and EV to a point on the frequency $\times$ component spectrum.

I previously state that DDK instrument with the common factor in Section 6 and allude to a later explanation. I now provide this explanation. DDK include stock-quarter FE and the regression therefore identifies coefficients using only the variation from within a stock-quarter. This absorbs all variation over ≈ 62 days. When comparing DDK to the coefficients from the frequency $\times$ component spectrum I focus on coefficients which correspond to periods of less than 62 days. Turning to the coefficients in the DDK model *dark* is estimated with a negative relation to *DEPTH* and a positive relation to *ES*, but when instrumented with itself (*dark*), the signs on both relations flip (Table A9). Whereas $FRAG^{lit}$only flips when instrumented with both *dark* and $FRAG^{lit}$(Table A9). While this disparity is initially puzzling, the frequency $\times$ component spectrum explains the sign flip. The coefficients estimating the relation between *dark* and both measures of liquidity corresponding to bands of less than 62 days are opposite in sign for 10/12 bands. This implies that for periods of less than 62 days, the idiosyncratic and common components are associated with opposite relations. For $FRAG^{lit}$the common and idiosyncratic components agree in sign for all 12 bands. When *dark* is instrumented with its cross-sectional average in DDK, the relation swaps from the idiosyncratic to the common, flipping the sign. When $FRAG^{lit}$is instrumented with the common component of itself, the sign remains constant.

The flip on *dark* but not $FRAG^{lit}$indicates that a leave-one-out group average as an instrument in 2SLS does not purge the regressor of bias but rather it replaces the regressor with a new object. Group averages of fragmentation variables do not appear to be a clean object of a firm's own fragmentation but rather an estimate of the common component of fragmentation. Whether the substitution of the object is damaging to the design is dependent on whether the idiosyncratic and common components share a direction in their relation to the outcome variable. Other microstructure studies employ this same instrument. The first reference I locate is Hasbrouck and Saar (2013) who use the group average IV but state "*There are no clear and obvious candidates for instrumental variables*" (p. 665). The paper also reports that the results strengthen under the IV and reason that: i) the results may suggest a prominent role of common determinants and ii) the 2SLS effectively removes noise from each stock's process. My findings here suggest that the IV plays the role of the former of the two statements. Excluding DDK, other papers have also used the same group average instrument to investigate dark fragmentation (Comerton-Forde and Putniņš, 2015; Buti, Rindi and Werner, 2022). All four of these papers also discuss the limitations of the group average IV but only Hasbrouck and Saar (2013) discuss the potential of the IV to stand in for a common effect.

Moving to discuss the final of the three estimators I employ, the EV specification strips memory with the first differences and is meant to capture the idiosyncratic relation as it projects out the fractionally integrated common factors. The EV setup therefore has a direct comparison to the frequency $\times$ component spectrum and is analogous to the shortest period idiosyncratic relation from Figure 7. The EV specification coefficients compared to the 2-3 day idiosyncratic period coefficients are: -0.2390 versus -0.234 for ES on $FRAG^{lit}$, 0.1355 versus 0.168 for ES on *dark*, 0.5629 versus 0.672 for $DEPTH$ on $FRAG^{lit}$, and -0.4701 versus -0.487 for $DEPTH$ on *dark*. Summarizing the three estimators I implement, the pooled specification presents one coefficient which represents an average relation across the frequency $\times$ component spectrum. The DDK specification instruments idiosyncratic variation with common variation. The EV specification returns the shortest period idiosyncratic relation.

I do not make the previous three comparisons to declare one of the models "correct," but rather to demonstrate that any model which reports a single coefficient inherently reports only one portion of the frequency $\times$ component spectrum, or as in the levels specification collapses the entire relation into one point estimate. I attempt to use my results to resolve conflicting models and contradictory past empirical findings within the fragmentation literature. I underscore that this is not a test, and put plainly, I have forty coefficients from the component $\times$ frequency spectrum which I have the freedom to now "connect" post-hoc to any existing paper's findings. I cannot create an all-encompassing synthesis of the extant fragmentation literature as my results do not permit that, but I am able to map ostensibly contradictory results onto my findings.

Parlour and Seppi (2003) model that competition can increase or decrease aggregate liquidity and nearly two decades later theoretical work has not asserted a direction, stating the relation is dependent on competing channels (Baldauf and Mollner, 2021). My results imply that both channels can be at play, living in different frequencies and components. At shorter periods the competition channel dominates and spreads tighten but as the overall market structure drifts towards increased lit fragmentation spreads widen. Consider the models of Pagano (1989); Chowdhry and Nanda (1991) where both predict equilibria whereby trading costs decrease with consolidation. These equilibria are long horizon state outcomes which map to my finding that long fragmented periods are associated with increased trading costs, evidenced by the positive relation above 125 days.

At shorter frequencies, my results suggest that for periods below 62 days, the association of lit fragmentation and spreads is negative. Empirical findings on competition reducing spreads in short time frames can be found in Degryse, De Jong and Kervel (2015) who attribute the tightening of spreads to competition between liquidity providers.[47] Baldauf and Mollner (2021) estimate a model where the number of venues is fixed and exchanges re-optimize fees within each regime. Their counterfactual monopoly spread is 22.9% below the observed duopoly spread. This counterfactual monopoly coincides with the relation estimate from the long period outcomes, whereby spreads are tightened with consolidation. Haslag and Ringgenberg (2023) report a coefficient that associates lit fragmentation with tighter spreads, but loses significance when including controls. Fragility may be a symptom of a single coefficient forced to describe a relation that spans a heterogeneous spectrum. Mapping the relation of fragmentation with aggregate depth is simpler. Depth is generally predicted to be increasing with competition (Baldauf and Mollner, 2021) and empirical findings coincide (Boehmer and Boehmer, 2003; Fink, Fink and Weston, 2006). This is consistent with my findings that lit fragmentation is associated with increased depth at all periods.

The mapping of previous literature to my findings can also be carried out in relation to off-exchange fragmentation. My findings suggest that idiosyncratic increases in off-exchange trading are associated with thinner depth and wider spreads at all periods. This is consistent with the theoretical model of Zhu (2014) who states that when all else is equal a dark pool accompanying an exchange will

[47] The mechanism hinted at is liquidity providers who bypass time priority as in Foucault and Menkveld (2008). I do not map the findings of Foucault and Menkveld (2008) directly to my period-dependent findings as the introduction of LSE into the Dutch markets, the object of study, was accompanied by a decrease in fees from the incumbent exchange.

push informed traders to the lit market and can contemporaneously widen spreads. Buti, Rindi and Werner (2017) model a dark pool competing with an LOB, causing an increase in spreads similar to Zhu (2014) and also a decrease in depth. The segmentation of order flow from Zhu (2014) is found empirically in Comerton-Forde and Putniņš (2015) where dark trading increases adverse selection in the lit market and the spread widens.[48] My study is in agreement as an increase in idiosyncratic dark trading is associated with a thinned book and widened spreads. Explaining the common component of the relation between dark fragmentation and liquidity is more difficult.

The literature has recently noted that the effects of dark trading are complex and it may be misleading to focus on one form, part of the cross-section or time period (Buti, Rindi and Werner, 2022). My study investigates how the relation of dark fragmentation and liquidity varies continuously across period and component. A horizon-dependent relation with liquidity is not a new proposition. Hendershott and Mendelson (2000) study crossing networks and find that the longevity of information determines the effect on spreads. Zhu (2014) introduces the information horizon as a parameter when modeling dark trading. Buti, Rindi and Werner (2017) note the effects of introducing a dark pool can vary over time. Dark and lit trading shares have also been shown to dynamically respond to information on the intraday level (Menkveld, Yueshen and Zhu, 2017). While the dynamic relation across the period domain of my study is not equivalent to the longevity of information as modeled in the cited works, I borrow the dynamic framework that decades of literature have established.

Initially puzzling is my finding that common dark fragmentation is associated with improved depth at all periods, and improved trading costs under 40 days yet deteriorated trading costs above 40 days. I do not interpret the results as the relation of common dark trading and liquidity "reversing" after 40 days but rather that common activity reflects equilibrium conditions pertaining to the allocations of order flow across lit and dark venues. At shorter periods the relation is consistent with endogenous venue selection under liquid, low-urgency conditions (Buti, Rindi and Werner, 2017; Menkveld, Yueshen and Zhu, 2017). Similarly, Zhu (2014) notes that competition or inventory costs may simultaneously increase dark pool participation and reduce lit spreads. At longer periods, uninformed order flow may persistently migrate off-exchange, increasing adverse selection on the lit exchange which widens the spreads and reduces the rate at which displayed orders are consumed. Consistent with this interpretation, Gresse (2017) finds that dark trading can be associated with greater depth and wider spreads and Buti, Rindi and Werner (2017) show that spread and depth responses are conditional on whether it is market or limit orders which migrate.[49]

[48] Further evidence is provided in Comerton-Forde, Malinova and Park (2018) where a regulation change shifts retail activity from off-exchange to the lit exchange and spreads improve.

[49] Boulatov and George (2013) model the ability of informed traders to hide their orders and find that when informed

Consequently, my period and component findings allow for the placement of several seemingly conflicting findings across the literature related to market fragmentation (whether lit or dark). It should be clear that not every paper related to market fragmentation will directly map to my empirical findings. Results may differ across markets, samples, and especially time frames, since I present that both idiosyncratic and common components of fragmentation are long memory and a study may randomly coincide with a drift in one direction if the sample is short. The purpose is to provide discussion of a seemingly conflicted literature. I am able to map numerous contradictory theoretical predictions and empirical findings to my frequency $\times$ component spectrum. Overall, my findings imply the relation of fragmentation and liquidity cannot be condensed into a single point coefficient.

8 Conclusion

I document that market fragmentation (lit or dark) and liquidity (trading costs or depth) on the stock-level are long memory processes, with median differencing parameters of 0.58 for lit fragmentation and 0.57 for dark fragmentation across 1,758 stocks. I also document persistence in the aggregate series of fragmentation and liquidity, where aggregate lit fragmentation is more persistent than the typical stock at $\hat{d} = 0.73$. The stock-level fragmentation series are reduced but remain persistent after projecting out common components following Ergemen and Velasco (2017). The median estimate for lit fragmentation falls from 0.582 to 0.500 after the projection. This implies that the stock-level fragmentation process contains both idiosyncratic and common persistence. I also find the persistence of fragmentation varies in the cross section. I perform a diagnostic to determine whether cointegration is present in the data. Both the stock- and aggregate-level series of liquidity and fragmentation differ in their order of integration. The residuals from regressions of liquidity on fragmentation remain more persistent which is evidence against cointegration between fragmentation and liquidity (Engle and Granger, 1987).

To address the non-stationarity of both the series and common components, I remove idiosyncratic memory and project persistent common factors out of the system following Pesaran (2006); Ergemen and Velasco (2017). I find that lit fragmentation is associated with improved liquidity in both dimensions (trading costs or depth) and dark fragmentation is associated with deteriorated liquidity in both dimensions. These single associations conceal a more nuanced relation between fragmentation and liquidity.

traders are forced to display their orders they shift from liquidity providers to liquidity demanders and market quality deteriorates.

It is well known that variance accumulates at the lowest frequencies of a long memory time series (Baillie, 1996). I therefore investigate the frequency domain of the relation between fragmentation and liquidity using a fast Fourier transform (FFT). I then further decompose the frequency domain into idiosyncratic and common components. Lit fragmentation is associated with tightened spreads at periods below 125 days and widened spreads at longer periods, as well as increased depth at all periods, for both components. Idiosyncratic dark fragmentation is associated with deteriorated liquidity for both dimensions. The common component of dark fragmentation is associated with improved trading costs at relatively shorter periods, deteriorated trading costs at relatively longer periods, and improved depth at all periods. These findings must be treated as relations. While the design includes stock fixed effects, heterogeneous loadings on persistent common factors, and memory removal from both components, it does not identify causal effects. I find the relation of fragmentation and liquidity is complex and cannot be expressed in a single point coefficient. Rather, it must be described at segments over a continuous frequency $\times$ component spectrum. Under this interpretation, findings that appear to contradict one another need not be in conflict. They may simply describe different regions of the same spectrum.

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies* 1, 3–40.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics* 83, 413–452.

Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance* 52, 975–1005.

Aquilina, M., Budish, E., O'Neill, P., 2022. Quantifying the high-frequency trading "arms race". *The Quarterly Journal of Economics* 137, 493–564.

Arnold, T., Hersch, P., Mulherin, J.H., Netter, J., 1999. Merging markets. *The Journal of Finance* 54, 1083–1107.

Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.

Baldauf, M., Mollner, J., 2021. Trading in fragmented markets. *Journal of Financial and Quantitative Analysis* 56, 93–121.

Bandi, F.M., Chaudhuri, S.E., Lo, A.W., Tamoni, A., 2021. Spectral factor models. *Journal of Financial Economics* 142, 214–238.

Bandi, F.M., Tamoni, A., 2023. Business-cycle consumption risk and asset prices. *Journal of Econometrics* 237, 105447.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance* 59, 1481–1509.

Barber, B.M., Huang, X., Jorion, P., Odean, T., Schwarz, C., 2024. A (sub) penny for your thoughts: Tracking retail investor activity in taq. *The Journal of Finance* 79, 2403–2427.

Battalio, R.H., 1997. Third market broker-dealers: Cost competitors or cream skimmers? *The Journal of Finance* 52, 341–352.

Battalio, R.H., Holden, C.W., Pierson, M., Shim, J.J., Wu, J., 2026. Latency and the look-ahead bias in trade and quote data. *Available at SSRN 5907665* .

Battalio, R.H., Jennings, R.H., Saglam, M., Wu, J., 2023. Difficulties in obtaining a representative sample of retail trades from public data sources. *Available at SSRN 4579159* .

Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do etfs increase volatility? *The journal of finance* 73, 2471–2535.

Bennett, P., Wei, L., 2006. Market structure, fragmentation, and market quality. *Journal of Financial Markets* 9, 49–78.

Bibinger, M., Reiß, M., 2014. Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics* 41, 23–50.

Blevins, J.R., 2026. Pyelw: Exact local whittle estimation for long memory time series in python .

Boehmer, B., Boehmer, E., 2003. Trading your neighbor's etfs: Competition or fragmentation? *Journal of Banking & Finance* 27, 1667–1703.

Boehmer, E., Jones, C.M., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. *The Journal of Finance* 76, 2249–2305.

Bollerslev, T., Jubinski, D., 1999. Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics* 17, 9–21.

Bouchaud, J.P., Farmer, J.D., Lillo, F., 2009. How markets slowly digest changes in supply and demand, in: Hens, T., Schenk-Hoppé, K.R. (Eds.), *Handbook of Financial Markets: Dynamics and Evolution*. North-Holland, Elsevier, pp. 57–160.

Boulatov, A., George, T.J., 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. *The Review of Financial Studies* 26, 2096–2137.

Breidt, F.J., Crato, N., De Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. *Journal of econometrics* 83, 325–348.

Bryzgalova, S., Huang, J., Julliard, C., 2025. Macro strikes back: Term structure of risk premia. *Available at SSRN* .

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics* 130, 1547–1621.

Buti, S., Rindi, B., Werner, I.M., 2017. Dark pool trading strategies, market quality and welfare. *Journal of Financial Economics* 124, 244–265.

Buti, S., Rindi, B., Werner, I.M., 2022. Diving into dark pools. *Financial Management* 51, 961–994.

Byers, D., Davidson, J., Peel, D., 1997. Modelling political popularity: an analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160, 471–485.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29, 238–249.

Carr, P., Madan, D., 1999. Option valuation using the fast fourier transform. *Journal of computational finance* 2, 61–73.

Carr, P., Wu, L., 2004. Time-changed lévy processes and option pricing. *Journal of Financial economics* 71, 113–141.

Cboe Global Markets, 2026. U.S. Equities Market Volume Summary. URL: https://ww2.cboe.com/us/equities/market_statistics/. snapshot accessed September 2, 2026; market data delayed at least 20 minutes.

Chen, Y., Eaton, G.W., Paye, B.S., 2018. Micro (structure) before macro? the predictive power of aggregate illiquidity for stock returns and economic activity. *Journal of Financial Economics* 130, 48–73.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. *Journal of Financial Economics* 56, 3–28.

Chowdhry, B., Nanda, V., 1991. Multimarket trading and market liquidity. *The Review of Financial Studies* 4, 483–511.

Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society* , 135–155.

Comerton-Forde, C., Malinova, K., Park, A., 2018. Regulating dark trading: Order flow segmentation and market quality. *Journal of Financial Economics* 130, 347–366.

Comerton-Forde, C., Putniņš, T.J., 2015. Dark trading and price discovery. *Journal of Financial Economics* 118, 70–92.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. *Mathematical finance* 8, 291–323.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.

Degryse, H., De Jong, F., Kervel, V.v., 2015. The impact of dark trading and visible fragmentation on market quality. *Review of Finance* 19, 1587–1622.

Dew-Becker, I., Giglio, S., 2016. Asset pricing in the frequency domain: theory and empirics. *The Review of Financial Studies* 29, 2029–2068.

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.

Engle, R.F., 1974. Band spectrum regression. *International Economic Review* 15, 1–11.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.

Epps, T.W., Epps, M.L., 1976. The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica: Journal of the Econometric Society* , 305–321.

Eraker, B., 2004. Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of finance* 59, 1367–1403.

Ergemen, Y.E., Velasco, C., 2017. Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *Journal of Econometrics* 196, 248–258.

Fink, J., Fink, K.E., Weston, J.P., 2006. Competition on the nasdaq and the growth of electronic communication networks. *Journal of Banking & Finance* 30, 2537–2559.

Finra, 2022. Market cap explained. URL: https://www.finra.org/investors/insights/market-cap. accessed January 16, 2026.

Firouzi, S., Wang, X., 2019. A comparative study of exchange rates and order flow based on wavelet transform coherence and cross wavelet transform. *Economic Modelling* 82, 42–56.

Foley, S., Putniņš, T.J., 2016. Should we be afraid of the dark? dark trading and market quality. *Journal of Financial Economics* 122, 456–481.

Foucault, T., Menkveld, A.J., 2008. Competition for order flow and smart order routing systems. *The Journal of Finance* 63, 119–158.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.

Glosten, L.R., 1994. Is the electronic open limit order book inevitable? *The Journal of Finance* 49, 1127–1161.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics* 14, 71–100.

Gould, M.D., Porter, M.A., Howison, S.D., 2016. The long memory of order flow in the foreign exchange spot market. *Market Microstructure and Liquidity* 2, 1650001.

Gradojevic, N., 2012. Frequency domain analysis of foreign exchange order flows. *Economics Letters* 115, 73–76.

Granger, C.W., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of Empirical Finance* 11, 399–421.

Granger, C.W., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.

Greene, M.T., Fielitz, B.D., 1977. Long-term dependence in common stock returns. *Journal of Financial Economics* 4, 339–349.

Gresse, C., 2017. Effects of lit and dark market fragmentation on liquidity. *Journal of Financial Markets* 35, 1–20.

Harvey, A.C., 2007. Long memory in stochastic volatility, in: *Forecasting volatility in the financial markets*. Elsevier, pp. 351–363.

Hasbrouck, J., 1991. Measuring the information content of stock trades. *The Journal of Finance* 46, 179–207.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. *The Journal of Finance* 50, 1175–1199.

Hasbrouck, J., Saar, G., 2013. Low-latency trading. *Journal of Financial Markets* 16, 646–679.

Hasbrouck, J., Seppi, D.J., 2001. Common factors in prices, order flows, and liquidity. *Journal of Financial Economics* 59, 383–411.

Haslag, P., Ringgenberg, M.C., 2023. The demise of the nyse and nasdaq: Market quality in the age of market fragmentation. *Journal of financial and quantitative analysis* 58, 2753–2782.

Hendershott, T., Jones, C.M., 2005. Island goes dark: Transparency, fragmentation, and regulation. *The Review of Financial Studies* 18, 743–793.

Hendershott, T., Jones, C.M., Menkveld, A.J., 2011. Does algorithmic trading improve liquidity? *The Journal of Finance* 66, 1–33.

Hendershott, T., Mendelson, H., 2000. Crossing networks and dealer markets: Competition and performance. *The Journal of Finance* 55, 2071–2115.

Holden, C.W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance* 69, 1747–1785.

Huberman, G., Halka, D., 2001. Systematic liquidity. *Journal of Financial Research* 24, 161–178.

Hurvich, C.M., Moulines, E., Soulier, P., 2005. Estimating long memory in volatility. *Econometrica* 73, 1283–1328.

Kamara, A., Lou, X., Sadka, R., 2008. The divergence of liquidity commonality in the cross-section of stocks. *Journal of Financial Economics* 89, 444–466.

Koch, A., Ruenzi, S., Starks, L., 2016. Commonality in liquidity: A demand-side explanation. *The Review of Financial Studies* 29, 1943–1974.

Korajczyk, R.A., Sadka, R., 2008. Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics* 87, 45–72.

Kwan, A., Masulis, R., McInish, T.H., 2015. Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics* 115, 330–348.

Lillo, F., Farmer, J.D., 2004. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics* 8.

Lo, A.W., 1991. Long-term memory in stock market prices. *Econometrica: Journal of the Econometric Society* , 1279–1313.

Lobato, I.N., Velasco, C., 2000. Long memory in stock-market trading volume. *Journal of Business & Economic Statistics* 18, 410–427.

Malliavin, P., Mancino, M.E., 2009. A fourier transform method for nonparametric estimation of multivariate volatility .

Menkveld, A.J., Yueshen, B.Z., Zhu, H., 2017. Shades of darkness: A pecking order of trading venues. *Journal of Financial Economics* 124, 503–534.

Müller, U.K., Watson, M.W., 2015. Low-frequency econometrics .

Nguyen, V., Van Ness, B.F., Van Ness, R.A., 2007. Short-and long-term effects of multimarket trading. *Financial Review* 42, 349–372.

NYSE, 2025. Daily TAQ client specification, version 4.1b. Intercontinental Exchange, Inc. Appendix F: Participant IDs; adds “G” for 24X National Exchange.

NYSE, 2026. Daily TAQ client specification, version 4.3. Intercontinental Exchange, Inc. Appendix F: Participant IDs.

Ohanissian, A., Russell, J.R., Tsay, R.S., 2008. True or spurious long memory? A new test. *Journal of Business & Economic Statistics* 26, 161–175.

O’Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? *Journal of Financial Economics* 100, 459–474.

Ortu, F., Tamoni, A., Tebaldi, C., 2013. Long-run risk and the persistence of consumption shocks. *The Review of Financial Studies* 26, 2876–2915.

Pagano, M., 1989. Trading volume and asset liquidity. *The Quarterly Journal of Economics* 104, 255–274.

Parlour, C.A., Seppi, D.J., 2003. Liquidity-based competition for order flow. *The Review of Financial Studies* 16, 301–343.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.

Pesaran, M.H., 2015. Testing weak cross-sectional dependence in large panels. *Econometric Reviews* 34, 1088–1116.

Pesaran, M.H., 2021. General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics* 60, 13–50.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 126–135.

Qu, Z., 2011. A test against spurious long memory. *Journal of Business & Economic Statistics* 29, 423–438.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* 23, 1630–1661.

Sato, A.H., 2008. Application of spectral methods for high-frequency financial data to quantifying states of market participants. *Physica A: Statistical Mechanics and its Applications* 387, 3960–3966.

Shimotsu, K., 2010. Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501–540.

Shimotsu, K., Phillips, P.C., 2005. Exact local whittle estimation of fractional integration. *The Annals of Statistics* 33, 1890–1933.

Singleton, K.J., 2001. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.

Staff of the Division of Trading and Markets, 2013. Equity Market Structure Literature Review, Part I: Market Fragmentation. Staff Literature Review. U.S. Securities and Exchange Commission. URL: https://www.sec.gov/marketstructure/research/fragmentation-lit-review-100713.pdf. accessed September 2, 2026.

Staff of the Office of Analytics and Research, Division of Trading and Markets, 2025. Trade-Through Roundtable Supporting Data. Staff Research Note. U.S. Securities and Exchange Commission. URL: https://www.sec.gov/files/trade-through-roundtable-supporting-data.pdf. revised September 12, 2025; accessed September 2, 2026.

Tauchen, G.E., Pitts, M., 1983. The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society* , 485–505.

U.S. Securities and Exchange Commission, 2005. Regulation NMS. Final Rule, Release No. 34-51808, File No. S7-10-04. URL: https://www.sec.gov/files/rules/final/34-51808.pdf. accessed September 2, 2026.

U.S. Securities and Exchange Commission, 2010. Concept release on equity market structure. URL: https://www.sec.gov/rules-regulations/2010/01/concept-release-equity-market-structure. accessed January 16, 2026.

U.S. Securities and Exchange Commission, 2019. Division of trading and markets: Background paper on the market structure for thinly traded securities. URL: https://www.sec.gov/files/rules/policy/2019/thinly-traded-securities-tm-background-paper.pdf. accessed August 18, 2026.

U.S. Securities and Exchange Commission, 2022a. Order Competition Rule. Proposed Rule, Release No. 34-96495, File No. S7-31-22. URL: https://www.sec.gov/files/rules/proposed/2022/34-96495.pdf. accessed September 2, 2026.

U.S. Securities and Exchange Commission, 2022b. Proposed Rule to Enhance Order Competition. Fact Sheet for Release No. 34-96495. URL: https://www.sec.gov/files/34-96495-fact-sheet.pdf. accessed September 2, 2026.

U.S. Securities and Exchange Commission, 2025. Roundtable on Rule 611 of Regulation NMS: Transcript. Austin, Texas, December 16, 2025. URL: https://www.sec.gov/files/rule-611-roundtable-transcript.pdf. opening remarks of Chairman Paul S. Atkins; accessed September 7, 2026.

Van Ness, B., Van Ness, R.A., Woods, D., 2026. Factors in the illiquidity of thinly traded securities. *Financial Review* .

Velasco, C., 1999. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.

Wu, L., Zhang, R., Dai, Y., 2026. Spectral volume models: Universal high-frequency periodicities in intraday trading activities. *Management Science* 72, 6652–6685.

Zhao, C., Chen, Y., Wu, L., Dai, Y., Chen, E., Wu, L., Zhang, R., 2025. High-frequency liquidity in the chinese stock market: Measurements, patterns, and determinants. *Pacific-Basin Finance Journal* 90, 102681.

Zhu, H., 2014. Do dark pools harm price discovery? *The Review of Financial Studies* 27, 747–789.

Figures

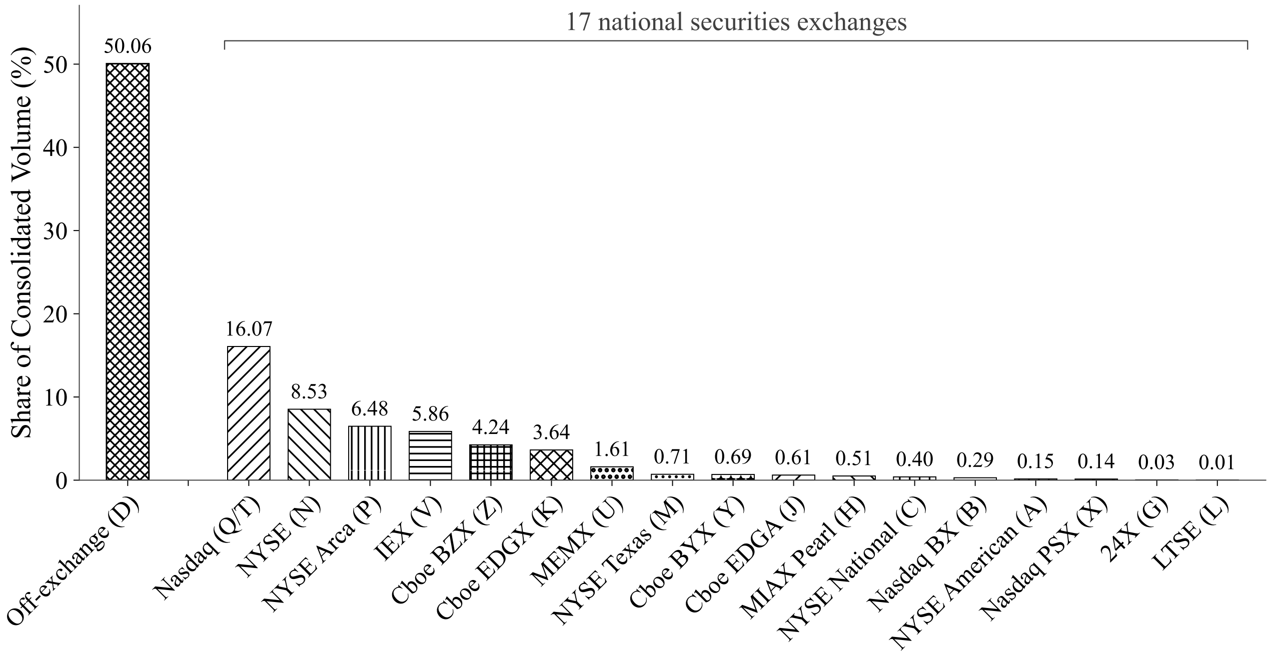
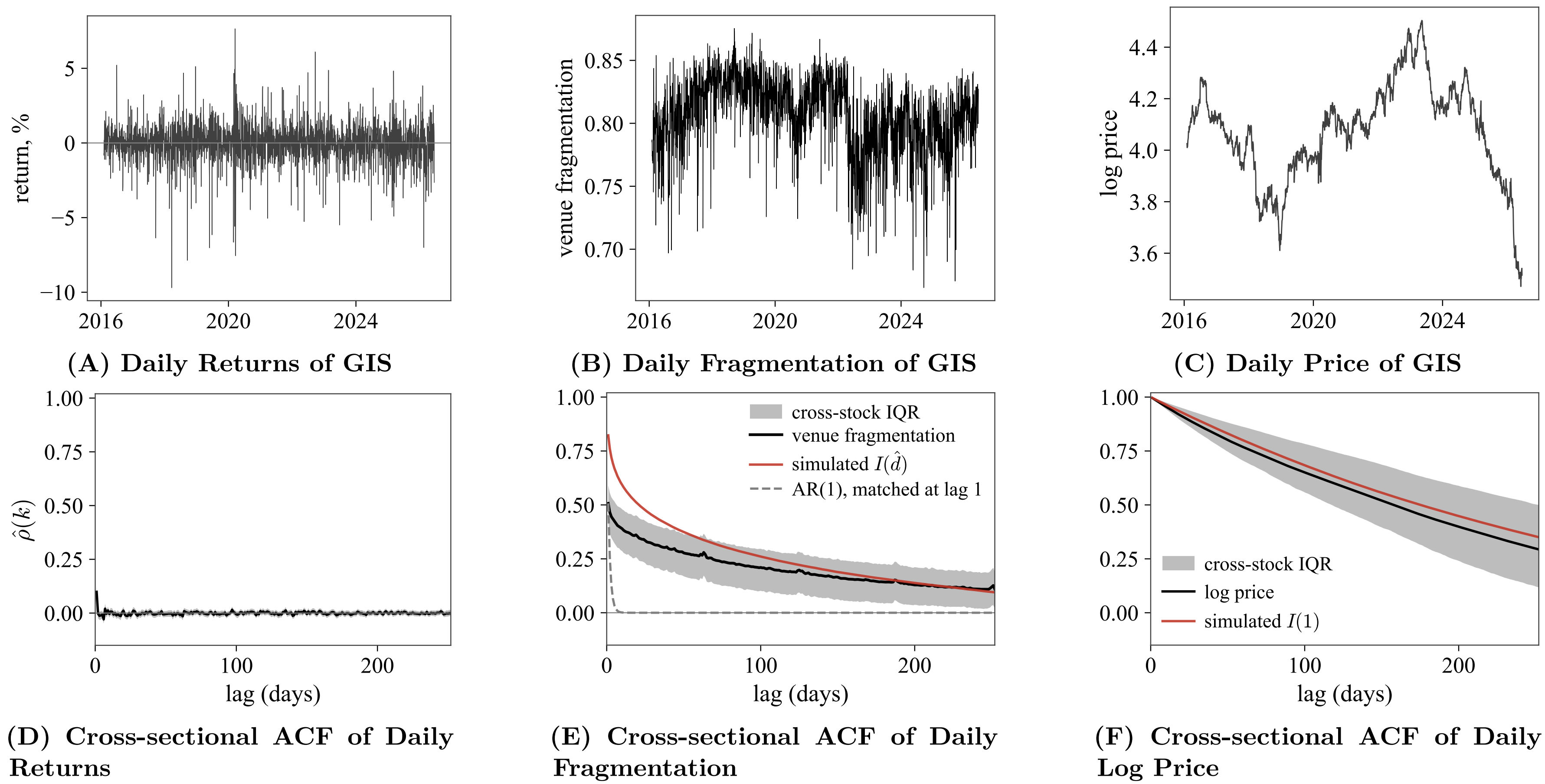


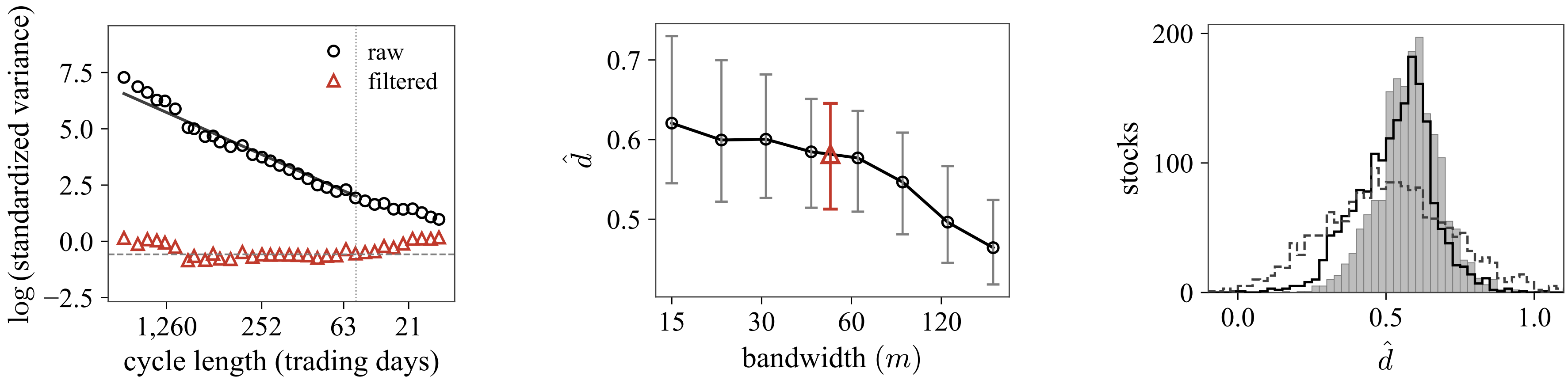
Figure 1: **Venue Shares of Consolidated Trading Volume.** The figure presents the share of consolidated trading volume executed at each exchange code in the TAQ data. The volume is pooled over the twenty trading days of May 2026 for the 1,758 stocks in the sample. Off-exchange is TAQ exchange code "D", the volume reported to a FINRA trade reporting facility. Each lit exchange is labelled with its TAQ participant code. Trades are restricted to regular trading hours, so the closing auction which prints after 16:00 is excluded.



(A) Daily Returns of GIS **(B) Daily Fragmentation of GIS** **(C) Daily Price of GIS**

(D) Cross-sectional ACF of Daily Returns **(E) Cross-sectional ACF of Daily Fragmentation** **(F) Cross-sectional ACF of Daily Log Price**

Figure 2: **Various Time Series** Panels (A)–(C) plot three daily series for General Mills (GIS), selected as the stock whose memory estimate is nearest the sample median. Panel (A) plots the daily returns, Panel (B) plots daily $FRAG^{lit}$and Panel (C) plots the daily log price. Panels (D)–(F) plot the cross-sectional mean sample autocorrelation functions for returns, $FRAG^{lit}$and log price across all 1,758 stocks, with the interquartile range shaded. In Panel (E) the $\hat{d}$ from each stock is entered into a simulated $\text{ARFIMA}(0, \hat{d}_i, 0)$ series matched to each stock's own series length, with the red line plotting the mean across stocks. The dashed line in Panel (E) is the population ACF of an AR(1) with $\phi = \hat{\rho}_1$ within each stock, equally weighted across stocks. The red line in Panel (F) is the simulated *sample* ACF of an I(1) process with length 2,609.

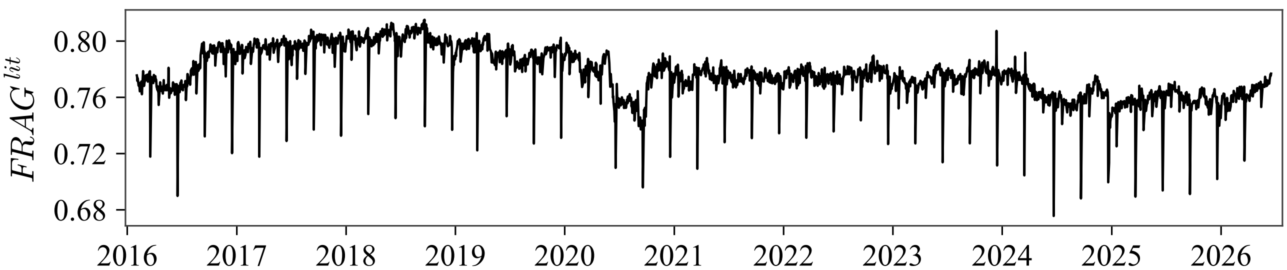


(A) Log Periodogram Pre/Post ELW

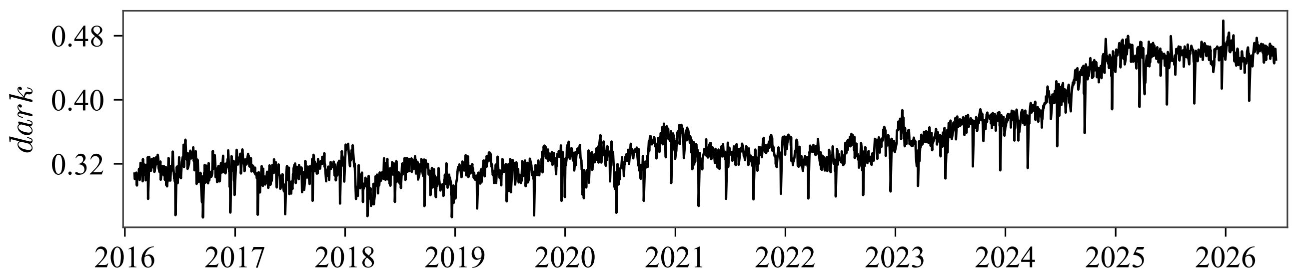
(B) $\hat{d}^{FRAG}$ Across m

(C) Cross-sectional Distribution of $\hat{d}^{FRAG}$

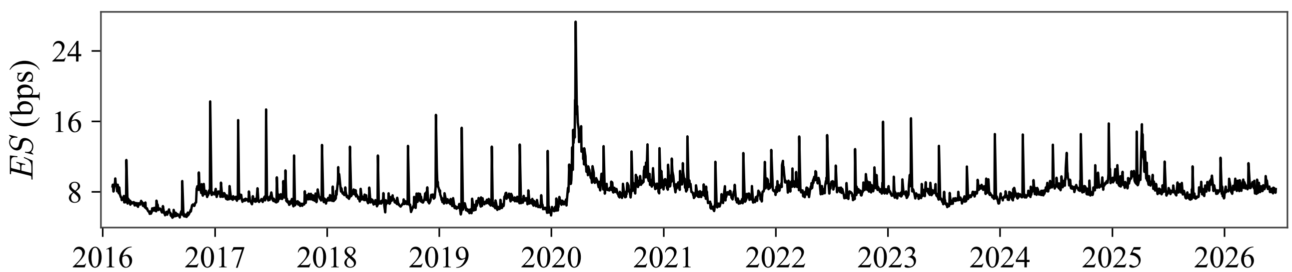
Figure 3: **Estimation of $\hat{\mathbf{d}}^{\mathbf{FRAG}}$** Panel (A) plots the log periodogram for $FRAG^{lit}$ where circles are the pooled periodogram of the raw series and triangles are the same series fractionally differenced by each stock's own estimate of d from exact local Whittle (Shimotsu, 2010). I scale the variance on the y-axis by $\hat{G}$, a scale parameter jointly estimated with d by ELW. The y-axis is not a proportion of variance. Exact local Whittle chooses d so that the differenced periodogram is flat across the bandwidth m selected. The base estimate of d uses a bandwidth of $m = n^{\frac{1}{2}}$ where n is the number of observations. The shortest period allowed into the bandwidth is denoted by the vertical dotted line. The horizontal dashed line represents what the log periodogram values would be under white noise. The solid line is a linear function with slope $-2\hat{d}$ at the median estimate across stocks. Panel (B) reports the stock median $\hat{d}$ against varying bandwidths m, with whiskers spanning the cross-stock interquartile range and the red triangle marking $m = n^{0.5}$. Panel (C) plots three cross-sectional distributions of per-stock $\hat{d}$: the raw estimates filled, the quadratically detrended ones as a solid outline, and the order-3 tapered estimates of Velasco (1999) as a dashed outline.



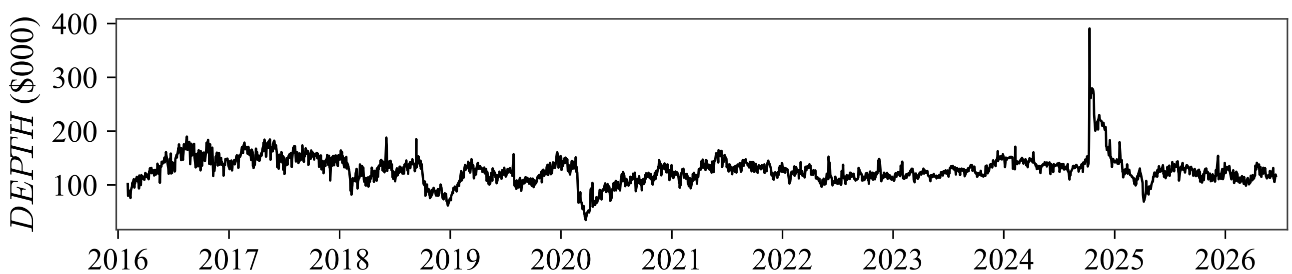
(A) Aggregate $\boldsymbol{FRAG^{lit}}$



(B) Aggregate $\boldsymbol{dark}$

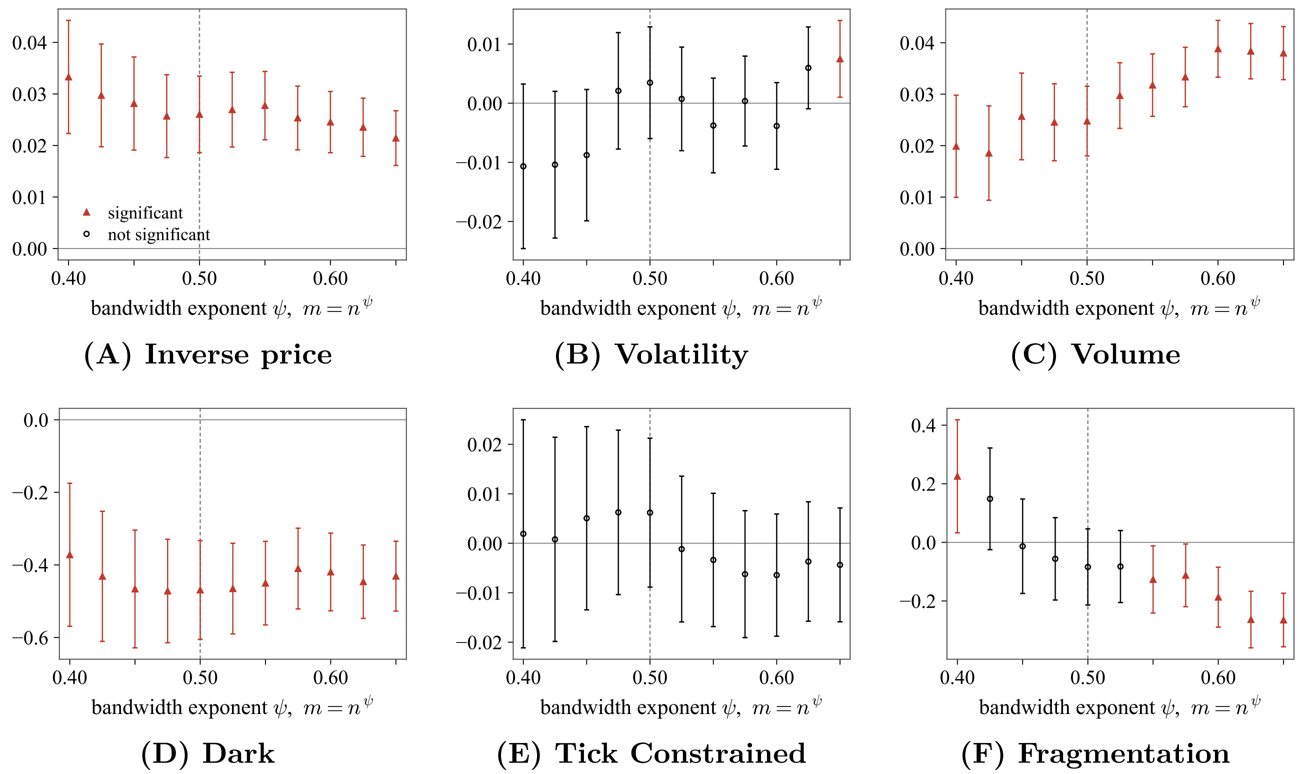


(C) Aggregate $\boldsymbol{ES}$



(D) Aggregate $\boldsymbol{DEPTH}$

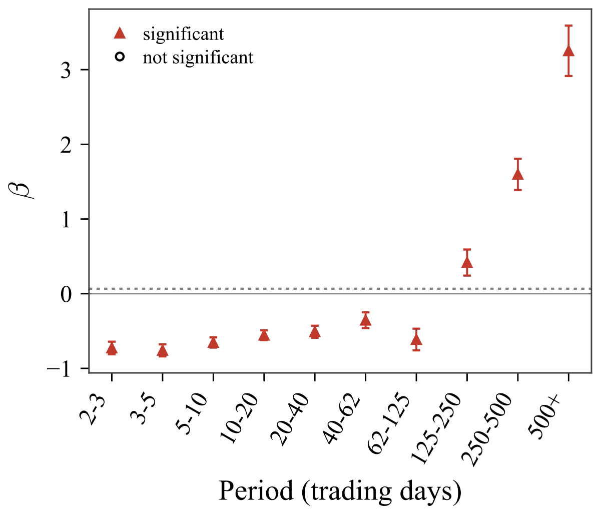
Figure 4: **The Aggregate Series** Each panel plots the equal-weighted cross-sectional mean of one daily series across every stock present on that day over 2,609 trading days. $FRAG^{lit}$ and $dark$ are plotted as shares, ES in basis points and $DEPTH$ in thousands of dollars.



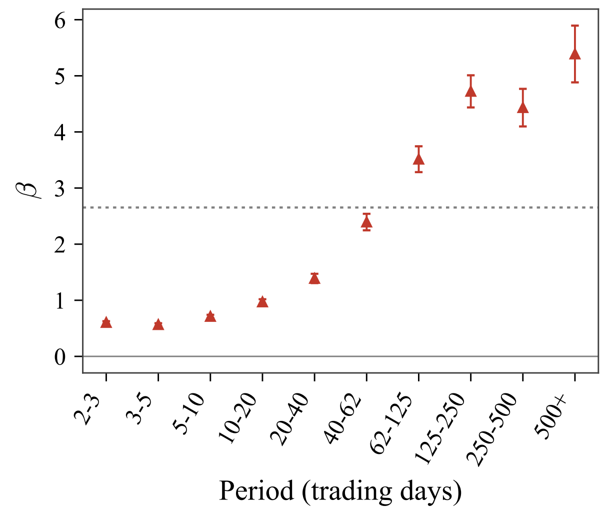
(A) Inverse price **(B) Volatility** **(C) Volume**

(D) Dark **(E) Tick Constrained** **(F) Fragmentation**

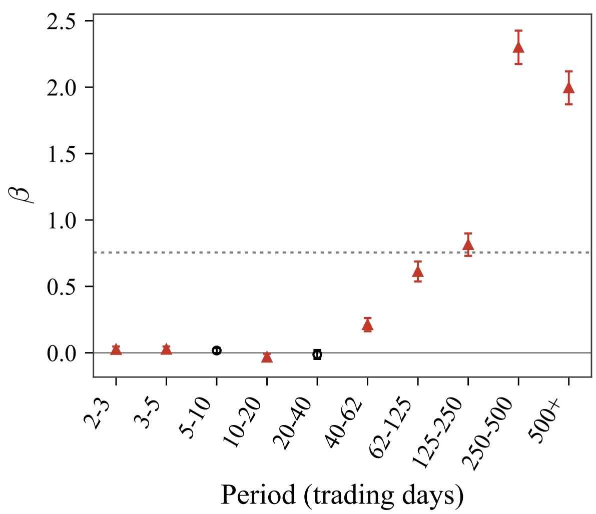
Figure 5: **Bandwidth sensitivity of the cross-section of persistence.** Each panel plots the coefficient on one stock characteristic from the cross-sectional regression of $\hat{d}^{FRAG}$ on all 6 characteristics jointly. The x-axis is reestimations separately at differing bandwidths $m = n^{\psi}$, $\psi \in [0.40, 0.65]$ (m from 23 to 166). $\hat{d}^{FRAG}$ is the ELW (Shimotsu, 2010) estimate for $FRAG^{lit}$. The dashed vertical rule marks the bandwidth $\psi = 1/2$ ($m = 51$) used throughout the paper. Whiskers are 95% confidence intervals from heteroskedasticity robust standard errors. Hollow black circles mark bandwidths at which the coefficient is not distinguishable from zero at the 5% level and solid red triangles mark those at which it is.



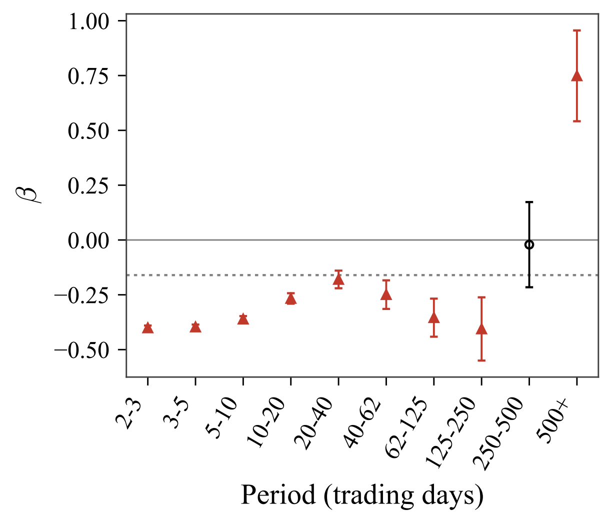
(A) *ES* on *FRAG*^{lit}



(B) *DEPTH* on *FRAG*^{lit}

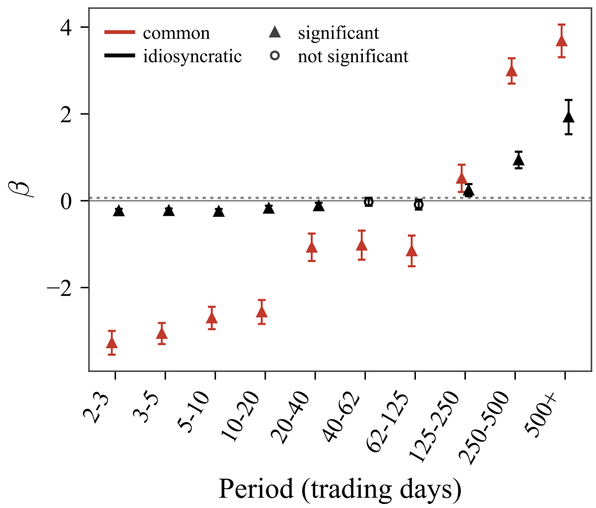


(C) *ES* on *dark*

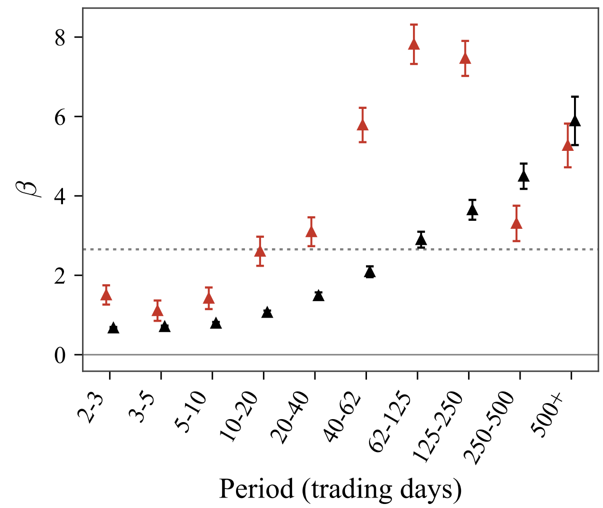


(D) *DEPTH* on *dark*

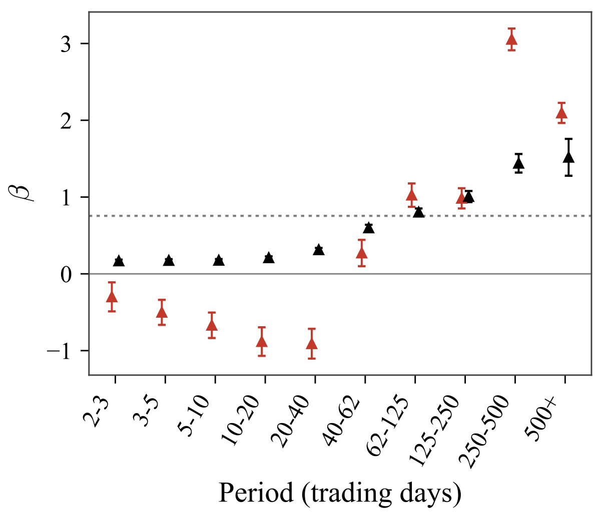
Figure 6: **Relation of Liquidity and Fragmentation Across Periods.** Each panel presents the coefficients from a panel regression of a liquidity measure, *ES* or *DEPTH*, on a fragmentation measure, $FRAG^{lit}$or *dark*, with control variables. Each regression is estimated on a panel that is fitted to the values of a pooled group of periods. The full ranges of the ten bands are [2-3), [3-5), [5-10), [10-20), [20-40), [40-62), [62-125), [125-250), [250-500), [500, 2,609) days. The regression estimated is $Y_{i,t}^{\mathcal{B}} = \beta_1^{\mathcal{B}} FRAG_{i,t}^{lit,\mathcal{B}} + \beta_2^{\mathcal{B}} dark_{i,t}^{\mathcal{B}} + \gamma^{\mathcal{B}\prime} Z_{i,t}^{\mathcal{B}} + \varepsilon_{i,t}^{\mathcal{B}}$ where $\mathcal{B} \in \{1, 2, \ldots, 10\}$ and $Z^{\mathcal{B}}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. $FRAG^{lit}$is 1 minus the HHI of trading volume shares using only lit exchanges and *dark* is the share of off-exchange volume. To compute the fitted panels which the regression is run on, first a FFT is run on each stock's time series, this yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \ldots, \frac{n}{2}\}$ and n is the number of daily observations for stock i. The periods are then pooled into the ten stated bands and OLS is estimated on the ten fitted bands. Each coefficient is from one of the twenty estimations, ten bands for each of the two liquidity measures. Standard errors are clustered on stock and date (Cameron, Gelbach and Miller, 2011). The FFT is mean zero by construction and takes the place of stock fixed effects. The dotted line presents the pooled coefficient, estimated on the same panel with the same control variables and no band-pass applied. Red triangles indicate significance at the 5% level and black circles are insignificant by the same standard. The whiskers display the 95% confidence intervals. The sample is 3,630,896 stock days over 1,672 stocks during 2,609 trading days.



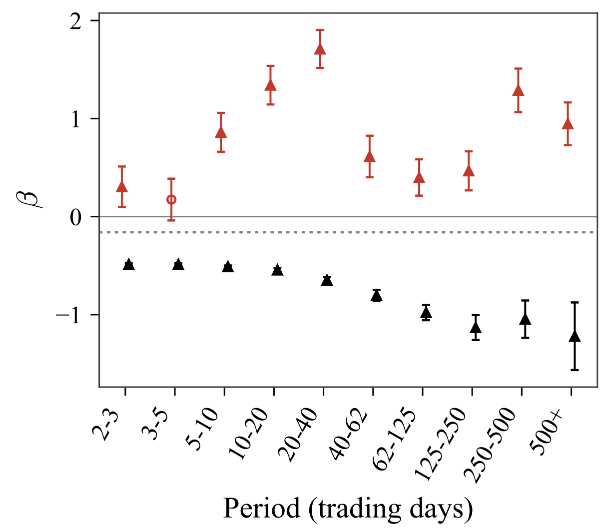
(A) $\boldsymbol{ES}$, $\boldsymbol{FRAG^{lit}}$



(B) $\boldsymbol{DEPTH}$, $\boldsymbol{FRAG^{lit}}$



(C) $\boldsymbol{ES}$, $\boldsymbol{dark}$



(D) $\boldsymbol{DEPTH}$, $\boldsymbol{dark}$

Figure 7: **Relation of Liquidity and Fragmentation Across Periods Decomposed by Components.** Each panel presents the coefficients from a panel regression of a liquidity measure, *ES* or *DEPTH*, on a fragmentation measure, $FRAG^{lit}$or *dark*, with control variables, decomposed between idiosyncratic and common component relations. Each regression is estimated on a panel that is fitted to the values of a pooled group of periods for one of the components, either common or idiosyncratic. The full ranges of the ten bands are [2-3), [3-5), [5-10), [10-20), [20-40), [40-62), [62-125), [125-250), [250-500), [500, 2,609) days. The regression estimated is $Y_{i,t}^{\mathcal{B},s} = \beta_1^{\mathcal{B},s} FRAG_{i,t}^{lit,\mathcal{B},s} + \beta_2^{\mathcal{B},s} dark_{i,t}^{\mathcal{B},s} + \gamma^{\mathcal{B},s\prime} Z_{i,t}^{\mathcal{B},s} + \varepsilon_{i,t}^{\mathcal{B},s}, \qquad s \in \{c, \iota\}$ where $\mathcal{B} \in \{1, 2, \ldots, 10\}$ and $Z^{\mathcal{B},s}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. $FRAG^{lit}$is 1 minus the HHI of trading volume shares using only lit exchanges and *dark* is the share of off-exchange volume. To compute the fitted panels which the regression is run on, first each stock is demeaned, then let z represent any variables time series. z is then decomposed into its idiosyncratic and common components as follows $\ddot{z}_{i,t} = a_i^z + \phi_i^{z\prime}\bar{h}_t + u_{i,t}^z, \qquad z_{i,t}^c = \hat{a}_i^z + \hat{\phi}_i^{z\prime}\bar{h}_t, \qquad z_{i,t}^\iota = \hat{u}_{i,t}^z$ where c denotes the common and ι denotes the idiosyncratic components. $\bar{h}_t$ is the vector of day t cross-sectional averages of every variable in the system: the liquidity measure, $FRAG^{lit}$, *dark*, the four control variables, and the leave-one-out size group averages of $FRAG^{lit}$, $FRAG^{lit}$squared and *dark*. Then a FFT is run on each stock's time series, this yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \ldots, \frac{n}{2}\}$ and n is the number of daily observations for stock i. The FFT is performed separately on the idiosyncratic and common component stock series. The periods are then pooled into the ten stated bands, separately for common and idiosyncratic components, and OLS is estimated on the ten fitted bands for each component. Each coefficient is from one of the forty estimations, ten bands for each component and each of the two liquidity measures. Standard errors are clustered on stock and date (Cameron, Gelbach and Miller, 2011). The FFT is mean zero by construction and takes the place of stock fixed effects. The dotted line presents the pooled coefficient on the undecomposed series, estimated on the same panel with the same control variables and no band-pass applied, so both components are read against one benchmark. Red markers indicate coefficients for the common series and black markers indicate coefficients for the idiosyncratic series. A triangle marker denotes significance at the 5% level and a circle denotes insignificance by the same standard. The whiskers display the 95% confidence intervals. The sample is 3,630,896 stock days over 1,672 stocks during 2,609 trading days.

Tables

Table 1: Summary statistics

The sample contains 1,758 common stocks over 2,609 trading days during the period 2016-02-02 to 2026-06-17 which contains 3,773,606 stock-days. Panels A–C are computed over all stock-days whereas Panel D is one observation per stock. $FRAG^{lit} = 1 - HHI^{Lit}$ and $FRAG^{all} = 1 - HHI^{All}$, where $FRAG^{lit}$ and $FRAG^{all}$ are the lit- and all-venue fragmentation proxies. *dark* is the off-exchange (TAQ venue "D") share of volume, computed as the exchange "D" volume on a stock-day divided by the total volume. *ES* is the volume-weighted LF effective spread in basis points, computed using lit trades only. *DEPTH* is the time-weighted sum of displayed dollar depth at the latency-free NBBO. Price is the median of NBBO midpoints. *volume* is daily share volume. Realized volatility is the square root of the sum of squared one-minute midpoint returns. Return volatility is the 20-day rolling standard deviation of daily log close-to-close returns. Both volatility measures are scaled by $\sqrt{252}$ and reported in annual percentage points. *size* is shares outstanding times price. Trading days per stock is the length of the series each stock. A stock is tick constrained if its time-weighted quoted spread is below \$0.011 on at least half its trading days, following Kwan, Masulis and McInish (2015).

	N	Mean	S.D.	P10	Median	P90
Panel A: Venue concentration and market structure						
$FRAG^{lit}$	3,773,606	0.7774	0.0582	0.6946	0.7906	0.8391
$FRAG^{all}$	3,773,606	0.7696	0.0584	0.6970	0.7787	0.8333
dark	3,773,606	0.3525	0.1027	0.2262	0.3447	0.4874
Panel B: Liquidity						
ES (bps)	3,773,606	7.96	8.25	2.18	5.40	16.12
DEPTH (\$000)	3,773,606	127.34	304.19	24.73	67.41	230.71
Panel C: Stock characteristics and controls						
Price (\$)	3,773,606	73.52	121.09	7.91	40.47	163.62
volume (millions)	3,773,606	2.66	5.34	0.33	1.10	5.65
Realized volatility (%, annualized)	3,773,606	28.50	22.90	12.38	23.05	50.62
Return volatility, 20-day (%, annualized)	3,773,606	36.73	25.15	15.01	29.87	66.16
size (\$bn)	3,773,606	27.54	119.14	0.99	6.08	53.16
Panel D: Stock level (N = stocks, not stock-days)						
Trading days per stock	1,758	2,168	562	1,246	2,608	2,609
Tick constrained (0/1)	1,758	0.268	0.443	0.000	0.000	1.000

Table 2: Estimating $\hat{\mathrm{d}}$ on Fragmentation and Liquidity Measures

Panel A presents ELW (Shimotsu, 2010) estimates of d for $FRAG^{lit}$ for the 1,758 common stocks. The number of estimates of d is 1,754 as for some stocks the estimator fails to converge. The second row applies a logit $= \log(h/(1-h))$ transformation before estimation. I follow Bollerslev and Jubinski (1999) and set bandwidth $m = n^{1/2}$. Panel B estimates the differencing parameters at varying bandwidths m. Panel C reports the same estimator on *volume*, ES, $DEPTH$, *dark*. Detrended and tapered estimates of d on both series are reported in Panel A of Table A1.

	N	Median	Mean	Q_1	Q_3	$\hat{d} > 0$ (%)
Panel A						
$FRAG^{lit}$	1,754	0.5816	0.5794	0.5127	0.6454	100.0
Logit $FRAG^{lit}$	1,754	0.5906	0.5878	0.5234	0.6527	100.0
Panel B						
$m = n^{0.4}$	1,737	0.5982	0.6088	0.5222	0.6953	100.0
$m = n^{1/2}$	1,754	0.5816	0.5794	0.5127	0.6454	100.0
$m = n^{0.6}$	1,754	0.5257	0.5273	0.4671	0.5927	100.0
Panel C						
Log *volume*	1,753	0.4981	0.4938	0.3943	0.5974	100.0
Log *ES*	1,750	0.6624	0.6855	0.5931	0.7729	100.0
Log *DEPTH*	1,754	0.7453	0.7557	0.6599	0.8476	100.0
dark	1,753	0.5747	0.5555	0.5017	0.6317	99.9

Table 3: The Common and the Idiosyncratic Component

The aggregate $\hat{d}$ reports the two-step ELW estimate of Shimotsu (2010) at $m = n^{1/2}$ on the equal-weighted cross-sectional average of each series. One series gives one estimate, so Days is the number of trading days. The stock-level $\hat{d}$ removes the common component before estimating following Pesaran (2006); Ergemen and Velasco (2017) where each stock's series is regressed on the day-t cross-sectional averages of all four series and a constant, and d is estimated on the residual using ELW. Stocks is the number of stocks the median is taken across, the observations vary as some stock-level series fail to converge using ELW. Gap is the aggregate $\hat{d}$ less the stock-level median, and Gap (s.e.) expresses it in units of the standard error of the aggregate estimate, $1/(2\sqrt{m})$ inflated by 1.17. The asymptotic value understates the finite-sample dispersion at a bandwidth this small, which is why I apply the 1.17 scaling Shimotsu (2010). The standard error is the same for every row.

	Days	Aggregate $\hat{d}$	Stocks	Stock-Level Median $\hat{d}$	Gap	Gap (s.e.)
$FRAG^{lit}$	2,609	0.7344	1,758	0.5003	0.2341	2.9
dark	2,609	0.7378	1,757	0.3679	0.3700	4.5
Log *ES*	2,609	0.6086	1,756	0.6091	-0.0005	0.0
Log *DEPTH*	2,609	0.7376	1,758	0.6782	0.0595	0.7

Table 4: Cointegration of Fragmentation and Liquidity

Panel A first runs a stock-level regression of $y = a + bx + \varepsilon$ for each pair denoted by $(y - x)$. The residuals ε are then collected and exact Local Whittle (Shimotsu, 2010) is estimated on the residual. The differencing parameter from ELW on the residual is denoted as $\hat{d}^{\varepsilon}$ and the median across stocks is tabulated. The reduction on the stock-level is then computed as $\min(\hat{d}^{x}, \hat{d}^{y}) - \hat{d}^{\varepsilon}$, and the median across stocks is tabulated. The final column of Panel A presents the % of stocks whose reduction is less than zero.
Panel B first runs an aggregate-level regression of $y = a + bx + \varepsilon$ for each pair denoted as $(y - x)$ and the residuals ε are extracted. Then ELW is estimated on the residuals at varying bandwidths m where $m \in \{0.4, 0.5, 0.6\}$. For each combination of series (y-x) and bandwidth m, the reduction $\min(\hat{d}^{x}, \hat{d}^{y}) - \hat{d}^{\varepsilon}$ is tabulated along with the reduction divided by the standard error in parentheses. The standard error is computed as $\frac{1}{2\sqrt{m}}$ (Shimotsu and Phillips, 2005).

Panel A: Stock level				
Pair $(y - x)$	N	$\hat{d}^{\varepsilon}$	Reduction	Stocks with reduction < 0 (%)
Log ES – $FRAG^{lit}$	1,750	0.6473	-0.0769	78.9
Log $DEPTH$ – $FRAG^{lit}$	1,754	0.7170	-0.1384	84.9
Log ES – $dark$	1,749	0.6392	-0.0830	71.7
Log $DEPTH$ – $dark$	1,753	0.7086	-0.1447	88.5
$dark$ – $FRAG^{lit}$	1,753	0.5615	-0.0045	60.3
$FRAG^{lit}$ – Log $volume$	1,753	0.5683	-0.0676	77.0
$dark$ – Log $volume$	1,752	0.5562	-0.0507	71.5

Panel B: Aggregate series, reduction $\min(\hat{d}^{x}, \hat{d}^{y}) - \hat{d}^{\varepsilon}$			
Pair $(y - x)$	$m = n^{0.4}$	$m = n^{1/2}$	$m = n^{0.6}$
Log ES – $FRAG^{lit}$	+0.1483 (+1.42)	+0.0097 (+0.14)	-0.0254 (-0.54)
Log $DEPTH$ – $FRAG^{lit}$	+0.0315 (+0.30)	-0.0262 (-0.37)	-0.1452 (-3.07)
Log ES – $dark$	+0.0486 (+0.47)	+0.0373 (+0.53)	-0.0199 (-0.42)
Log $DEPTH$ – $dark$	-0.0002 (-0.00)	-0.0132 (-0.19)	-0.1714 (-3.63)
$dark$ – $FRAG^{lit}$	+0.0604 (+0.58)	+0.0820 (+1.17)	+0.0522 (+1.11)
$FRAG^{lit}$ – Log $volume$	-0.1115 (-1.07)	-0.1327 (-1.90)	-0.2103 (-4.45)
$dark$ – Log $volume$	-0.1244 (-1.19)	-0.1146 (-1.64)	-0.1689 (-3.58)

Table 5: Cross-sectional Regressions of Persistence in $FRAG^{lit}$

The table presents cross-sectional regressions of the stock-level memory parameter $\hat{d}^{FRAG}$ on stock characteristics using one observation per stock. $\hat{d}^{FRAG}$ is the Shimotsu (2010) two-step ELW estimate. Characteristics are stock means over the sample window. The characteristics are: *inverse price*, the inverse of price, in log units; *volatility*, computed as the sum of squared minute-level midpoint returns, in log units; *size*, measured as the shares outstanding multiplied by the median midpoint of that day, in log units; *volume*, the daily share volume, in log units; *dark*, the share of volume off-exchange; *tick constrained*, an indicator for tick constrained stocks, constructed following the definition of Kwan, Masulis and McInish (2015); and *fragmentation*, the raw fragmentation level. The estimates in columns (1)-(9) use $m = n^{1/2}$ and column (10) uses $m = n^{0.4}$. VIFs for the specification in (8) are displayed in (11) and VIFs for the specification in (9) are presented in (12). Heteroskedasticity robust t-statistics in parentheses; *, ** and *** denote significance at the 10%, 5% and 1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Univariate							*Full*	*Reduced*	*Reduced*	*VIF*	
								$n^{1/2}$	$n^{1/2}$	$n^{0.4}$	Full	Reduced
Inverse price	0.0257***							0.0091	0.0261***	0.0333***	7.8	2.9
	(10.80)							(1.48)	(6.90)	(5.95)		
Volatility		0.0142***						-0.0076	0.0045	-0.0107	4.0	2.6
		(4.53)						(-1.25)	(0.94)	(-1.50)		
Size			-0.0090***					-0.0200***			12.2	
			(-4.89)					(-3.39)				
Volume				0.0201***				0.0433***	0.0259***	0.0199***	5.5	1.8
				(7.70)				(7.13)	(7.49)	(3.92)		
Dark share					-0.0834			-0.4960***	-0.4857***	-0.3721***	1.9	1.9
					(-1.61)			(-7.16)	(-7.00)	(-3.70)		
Tick constrained						0.0587***		0.0074	0.0056	0.0019	2.2	2.2
						(10.93)		(0.96)	(0.74)	(0.16)		
Fragmentation							0.3004***	-0.1119*	-0.0912	0.2256**	1.3	1.3
							(4.93)	(-1.66)	(-1.37)	(2.29)		
R^2	0.068	0.012	0.015	0.031	0.002	0.058	0.013	0.130	0.124	0.073		
N	1,754	1,754	1,754	1,754	1,754	1,754	1,754	1,754	1,754	1,737		

Table 6: Which fragmentation index measures venue competition?

The table presents the coefficients from panel regressions of liquidity outcomes on varying fragmentation proxies with stock and day fixed effects. The outcome variables are ES and $DEPTH$ and regressions are estimated on the contemporaneous and one day forward outcomes. ES is the volume weighted daily effective spread using the LF-NBBO (Battalio et al., 2026), in log units. $DEPTH$ is the summation of the time weighted depth at the LF best bid and LF best ask. $FRAG^{ALL}$ is one minus the HHI of exchange venue shares using all exchanges, $FRAG^{LIT}$ is one minus the HHI of exchange venue shares using only lit exchanges. $dark$ is the share of off-exchange trading volume to total volume. Control variables include: the daily share volume, in log units; the daily inverse price, in log units; the 20-day rolling standard deviation of daily log returns, in log units; the daily market capitalization, in log units. All variables are winsorized at the 0.5/99.5 percentile. Standard errors are two-way clustered on stock and date, t-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days spanning February 2, 2016 and June 17, 2026.

	Contemporaneous				Outcome at $t+1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Dependent variable: ES								
$FRAG_{\text{all}}$	-1.2607***	0.0796			-1.0621***	0.3701***		
	(-28.94)	(1.03)			(-24.85)	(5.81)		
$FRAG_{\text{lit}}$			0.6377***	0.5310***			0.8323***	0.7340***
			(9.76)	(8.36)			(14.25)	(13.01)
$dark$		0.9908***		0.9377***		1.0588***		0.8642***
		(20.55)		(35.35)		(25.57)		(33.13)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock and day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Within R^2	0.168	0.178	0.157	0.180	0.162	0.173	0.156	0.176
N	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606
Panel B. Dependent variable: $DEPTH$								
$FRAG_{\text{all}}$	1.5443***	2.2720***			1.1377***	2.0069***		
	(21.87)	(20.76)			(16.37)	(18.64)		
$FRAG_{\text{lit}}$			2.8566***	2.9259***			2.6107***	2.6531***
			(31.71)	(32.71)			(29.12)	(29.76)
$dark$		0.5379***		-0.6089***		0.6426***		-0.3725***
		(8.35)		(-15.54)		(10.11)		(-9.57)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock and day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Within R^2	0.204	0.206	0.230	0.238	0.186	0.189	0.213	0.216
N	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606	3,773,606

Table 7: Estimations from Ergemen and Velasco (2017) and Degryse, De Jong and Kervel (2015) Specifications

The table presents the coefficient estimates from regressions of liquidity variables ES and $DEPTH$ on fragmentation variables $FRAG^{lit}$and *dark*. ES is the volume weighted daily effective spread computed using the Battalio et al. (2026) LF-NBBO and signing methodology, and $DEPTH$ is time weighted sum of the depth at the best LF bid and ask price. Both liquidity outcome variables are in log units. $FRAG^{lit}$is 1 minus the HHI of exchange venue volume shares, using only lit exchanges. *dark* is the volume share of off-exchange volume. The control variables used in Ergemen and Velasco (2017) (EV) are: the daily share volume, denoted *volume*, in log units; the inverse price, denoted *price* in log units; the daily volatility, computed as the rolling 20-day standard deviation of log returns, denoted *volatility*, in log units; and the market capitalization, denoted *size* in log units. The steps of EV are as follows 1. all variables in the system are first differenced 2. all first differenced variables in the system are regressed on an intercept and a vector $\bar{H}_t$ which contains the day t cross-sectional averages of all variables *at the stock level* so all stocks are allowed heterogeneous loadings on the common factors 3. the residuals from the stock level regressions are pooled and the coefficients are estimated. The specification can be modeled by $\Delta Y_{i,t} = a_i + \beta_1 \Delta FRAG^{lit}_{i,t} + \beta_2 \Delta dark_{i,t} + \gamma' \Delta Z_{i,t} + \phi_i' \bar{H}_t + \varepsilon_{i,t}$. Standard errors are clustered two ways, on stock and on date (Cameron, Gelbach and Miller, 2011). The projection removes the average factor loading and not its dispersion, so the scores still co-move within a day; clustering on stock alone understates the standard error by 6% to 26% on these coefficients. Table A8 presents the full EV specification without controls truncated. Degryse, De Jong and Kervel (2015) (DDK) includes a squared term $(FRAG^{lit})^2$ and instruments the three endogenous variables $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark* using three instruments which are leave-one-out quartile size group averages of $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark*. The system of equations can be represented as (1) $FRAG^{lit}_{i,t} = a_{1,i}\delta_{q(t)} + b_1' X_{i,t} + c_1' W_{i,t} + \epsilon_{1,i,t}$, (2) $(FRAG^{lit}_{i,t})^2 = a_{2,i}\delta_{q(t)} + b_2' X_{i,t} + c_2' W_{i,t} + \epsilon_{2,i,t}$, (3) $dark_{i,t} = a_{3,i}\delta_{q(t)} + b_3' X_{i,t} + c_3' W_{i,t} + \epsilon_{3,i,t}$. The regression then estimates $Y_{i,t} = \alpha_i \delta_{q(t)} + \beta_1 \widehat{FRAG^{lit}}_{i,t} + \beta_2 (\widehat{FRAG^{lit}})^2{}_{i,t} + \beta_3 \widehat{dark}_{i,t} + \gamma' W_{i,t} + e_{i,t}$. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as $AvgLiq$; market cap, in log units, denoted as $Size$; the inverse price, in log units, denoted as $InversePrice$; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable $Algo$ in DDK's specification is standardized message traffic, as I do not have message data I proxy with NBBO updates divided by trading volume but note the limitation. $\alpha_i \delta_{q(t)}$ are stock-quarter fixed effects. DDK use Newey-West standard errors, but I use stock clustered. I report the marginal effect at the sample mean as well as the 10^{th} and 90^{th} percentiles. All variables in the DDK system are fractionally differenced at each stock's estimate of $\hat{d}^{FRAG}$, after differencing the first 250 observations of each stock's series are removed as they do not have significant pre-differencing observations to account for the long memory. The fractional differencing is applied before the stock-quarter demeaning. The system of equations is solved using GMM. Full tables for the DDK specification without differencing, without the quadratic specification and with controls not truncated are in the online appendix Table A6, and Table A7. t-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days.

	Ergemen–Velasco		Degryse et al., linear		Degryse et al., quadratic	
	ES	$DEPTH$	ES	$DEPTH$	ES	$DEPTH$
	(1)	(2)	(3)	(4)	(5)	(6)
$FRAG_{\text{lit}}$	-0.2390***	0.5629***	0.3173**	-0.5307***	-0.0900	-0.4294***
	(-10.87)	(55.35)	(2.32)	(-17.06)	(-0.51)	(-11.14)
at the 10th percentile					0.5084***	-0.6019***
					(3.33)	(-16.72)
at the 90th percentile					-0.5364**	-0.3006***
					(-2.05)	(-5.05)
dark	0.1355***	-0.4701***	-0.7368***	0.7828***	-0.7503***	0.8035***
	(17.45)	(-99.50)	(-18.27)	(27.81)	(-18.32)	(28.20)
Controls	Yes		Yes		Yes	
Fragmentation enters	linearly		linearly		quadratically	
Common factors removed	Yes		No		No	
Fixed effects	stock (differenced)		stock-quarter		stock-quarter	
Standard Errors	Stock and Date		Stock		Stock	
First-stage F, *VisFrag*			2,488	2,357	42	42
R^2	0.076	0.141	0.068	0.814	0.060	0.814
N	3,757,425	3,757,425	3,334,106	3,334,106	3,334,106	3,334,106

Online Appendix

The Persistence of Market Fragmentation

Michael Coccia

A Trends, Level Shifts, and Breaks

Estimations of long memory may be spurious in the presence of trends, level shifts and breaks within the time series (Velasco, 1999; Diebold and Inoue, 2001; Granger and Hyung, 2004; Shimotsu, 2010). I perform a battery of tests to assess whether the long memory estimates are indeed capturing true long memory. To demonstrate the necessity to implement these tests, in Table A2 I simulate true long memory in Panel A along with white noise contaminated with a linear trend or level shifts in Panel B. The table shows that spurious long memory may be identified under simple detrending, and therefore justifies the more aggressive level shifts and break detection I implement in this section.

To account for trends within the data I detrend the raw series and the logit series by linear ($k = 1$) and quadratic ($k = 2$) polynomials. The process for detrending involves regressing the stock-level series on the polynomial, and passing the residuals to ELW. The detrended estimates are presented in Panel A of Table A1; the median estimates of $\hat{d}$ are 0.5853 and 0.5592 for the linear and quadratic detrended raw $FRAG^{lit}$series respectively. While the detrended estimates are lower than the raw, they still confirm long memory in the non-stationary region. Detrending has two flaws: first, you must specify the order of the polynomial; and second, true long memory may present itself as a polynomial trend. An alternative is to use a tapered estimator which is independent of the order of the unobserved trend and does not remove any true long memory. I use the tapering procedure advanced by Velasco (1999) which is a third-order tapered local Whittle estimator. Tapering can be thought of as a pre-filtering of the data which smooths the abrupt start and end points of the series, which can be a source of spurious long memory. Nearly any time series possesses an abrupt start and stop point as it is a byproduct of sampling. A tapered third-order estimate is also invariant to linear and quadratic trends by construction; for a full discussion of tapering see Velasco (1999). The tapered estimate on $FRAG^{lit}$is 0.5166; while less than the raw and detrended estimates, it confirms long memory.

While I have demonstrated that unobserved trends are not the source of my long memory estimates, regime switching, which is referred to as structural breaks, may also cause spurious long memory detection (Diebold and Inoue, 2001). Granger and Hyung (2004) show that occasional breaks generate slowly decaying autocorrelations and the other observable properties of an $I(d)$ process. I emphasize that this concern is relevant for my setting and not a statistical curiosity. Fragmentation events such as exchange entrances (MEMX, MIAX, IEX), listing exchange changes, or cross-listing events within a stock, could be hypothesized to manufacture a level-shift in fragmentation. To test for the presence of structural breaks I use the W test of Qu (2011), which takes stationary long memory as the null and a process influenced by a regime change as the alternative. The W test is semiparametric and does not require specifying the form of the trend or the number of level-shifts. The W test rejects for 43.5% of stocks at the 5% level, against a size of 2.7% on simulated long-memory series with no breaks and a rejection rate of 91.0% on pure level shifts, each rate taken over 300 simulated series.[50] While 43.5% may appear damning initially, a rejection is not evidence against long memory; rather, it indicates the series contains a spurious component and that shifts or trends should enter the model, not that the memory itself is spurious (Qu, 2011). Indeed, one could roughly interpret the results of the test as describing the persistence of $FRAG^{lit}$somewhere in the middle between true long memory and a non-long memory process with breaks and level shifts.

I perform alternative tests to examine whether structural breaks drive the persistence in $FRAG^{lit}$, one of which involves splitting the sample and one which involves aggregating observations. True long memory is invariant to the frequency of observation (Ohanissian, Russell and Tsay, 2008), and a true long-memory process has a single d that does not depend on which portion of the sample is observed whereas a process of occasional breaks has neither property (Granger and Hyung, 2004).

For the 1,011 stocks with at least 2,400 daily observations, I split the time series into three windows, one which uses the first two years of data (2016–2018), one which uses the last two years of data (2024–2026) and one which uses the entire sample. $\hat{d}$ is estimated on each window and I take the difference between the full window and the first two-year window, and the difference between the full window and the last two-year window. Under true long memory, these differences should be zero or near zero. The results are presented in Panel C of Table A1. Along with the difference on the real data, I report the same difference on 300 simulated ARFIMA$(0, \hat{d}, 0)$ series with no breaks and on 300 simulated pure level shift series with no long memory, analogous simulations to Panel A, each simulated column being the mean paired change across its 300 replications. The differences for the real data are 0.030 and 0.150 for the first and last two-year windows respectively. The final column "Position" (%) of the table reports the position of the real data between the two simulated benchmarks, which is 6% and 31% for the first and last two-year windows respectively. Again, it appears that the observed time series of $FRAG^{lit}$across stocks contains long memory, but also some breaks and level shifts, which is not surprising.

The next alternative test I use involves summing daily observations into sequential eight-day observations and re-estimating $\hat{d}$ on the aggregated series.[51] The difference is reported in Panel C of

[50]The long-memory simulated series is ARFIMA$(0, \hat{d}, 0)$, the median ELW estimate. The level-shift benchmark is i.i.d./ standard normal noise plus a step function with three break dates drawn without replacement from the middle 80% of the sample; each break date moves the level up or down one σ. Series are generated at 2,608 observations, the panel's median length.

[51]Aggregation levels must be consecutive powers of two, and eight days is the coarsest level at which the size of the sample supports convergence in ELW.

Table A1 in the final row. As before, I include the same simulated benchmarks. The difference for the real data is 0.082 which lies 27% of the way between simulated pure long memory and the pure level shift process. Overall, the results of Panel C of Table A1 suggest that the long memory estimates of market fragmentation sit somewhere between 6% and 31% of the way from pure long memory toward pure level shifts. The positions are descriptive in nature, as they are purely an interpolation between two simulated benchmarks rather than a test with a standard error.

Qu (2011) recommends that after a rejection of pure long memory, one should remove the level shifts and reestimate the memory parameter. I remove level shifts using the Pettitt (1979) test. I perform this test both on stock-level $FRAG^{lit}$series, and the same simulated pure long memory series with no breaks at all. The results are presented in Panel D of Table A1. Interestingly, the Pettitt test detects more level shifts in the simulated series than in the real data. After removing the largest detected level shift, the re-estimated $\hat{d}$ decreases slightly for both the simulated and the real data. After removing all detected level shifts, $\hat{d}$ is re-estimated once again and both the simulated and real data $\hat{d}$ estimates are significantly lower than their respective original estimates. I emphasize that the differencing parameter estimated on the true long memory series is lower than that of the real fragmentation data. The results of Table A1 taken in tandem suggest that the long memory estimates of market fragmentation are true, but the data generating process of market fragmentation likely contains level shifts along with short-run dynamics not estimated by the semiparametric ELW estimator. I note that fitting an ARFIMA order by AIC selects an MA(1) for 99% of stocks.[52]

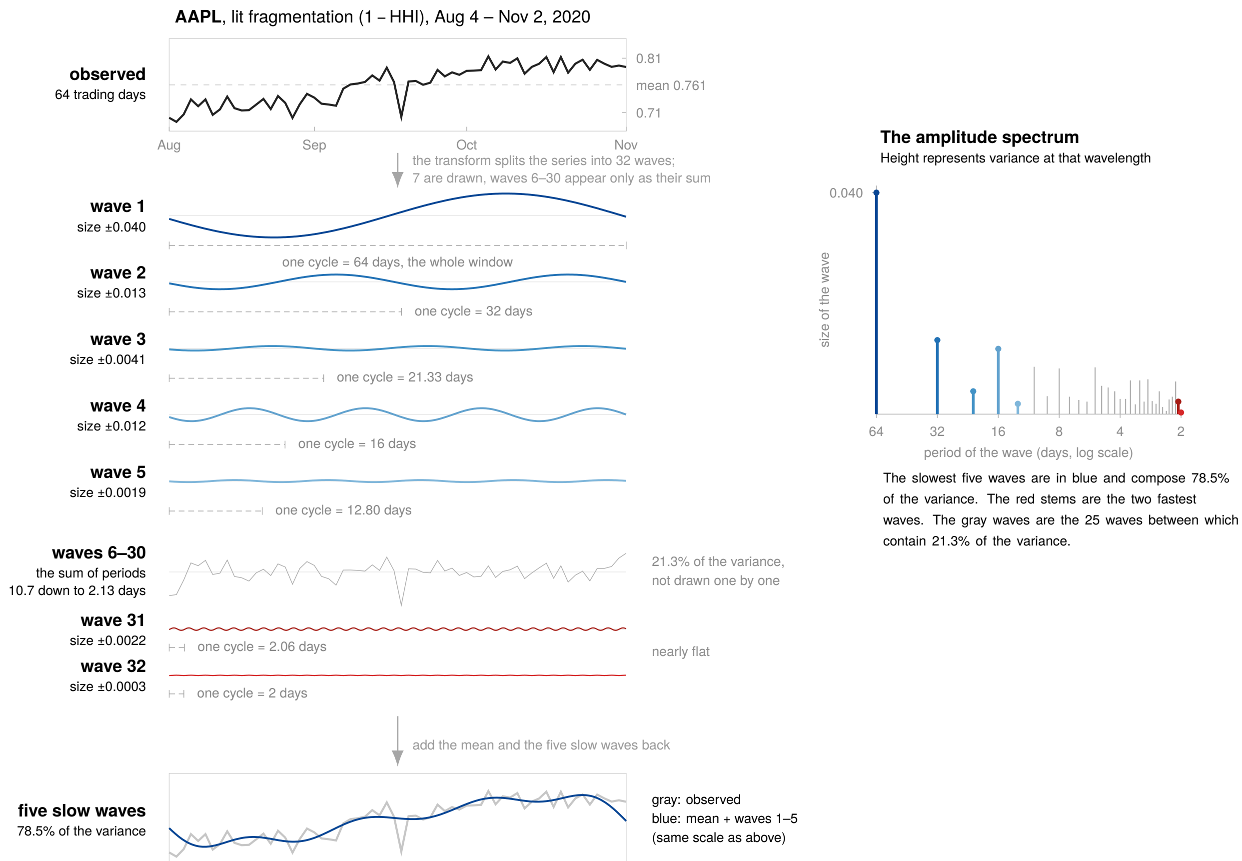
[52]The order is selected by AIC over $p, q \in \{0, 1\}$ in the pre-whitening step of Qu (2011), which is applied before the W test to correct its finite-sample size, not after the removal of level shifts.

Table A1: Long memory against level shifts: every statistic beside the benchmarks that make it readable

Panel A reports ELW (Shimotsu, 2010)at $m = n^{1/2}$, on $FRAG^{lit}$ and on its logit, after removing a polynomial trend in t of the stated order and after tapering. Detrended rows regress the series on the polynomial and estimate on the residual (Shimotsu, 2010) and tapered rows use the order-3 tapered local Whittle estimator of Velasco (1999). Panels B–d each report statistics, but use three different time series. The first is simulated long memory with no breaks. The second is fragmentation time series for the real sample and the third is a series simulated to be pure level shifts with no memory. The simulated columns are based on the panel's own length and, in the long-memory column, at its own $\hat{d}$. The Position column interpolates where the real series lies. It is defined as 100× (data − long memory) / (level shifts − long memory). Panel B applies the W test of Qu (2011) stock by stock, with his Section 5 pre-whitening and his tabulated critical value of 1.155 at trimming $\varepsilon = 0.05$. In Panel C the window gradient is the difference of the mean ELW estimate from the full sample and the first and last two years of the sample respectively. The third row follows Ohanissian, Russell and Tsay (2008) and $\hat{d}$ is re-estimated at 1 and 8 day aggregation, and the median change in magnitude is tabulated. In Panel D I detect level shifts using Pettitt (1979) and demean within the detected segments, and re-estimate $\hat{d}$.

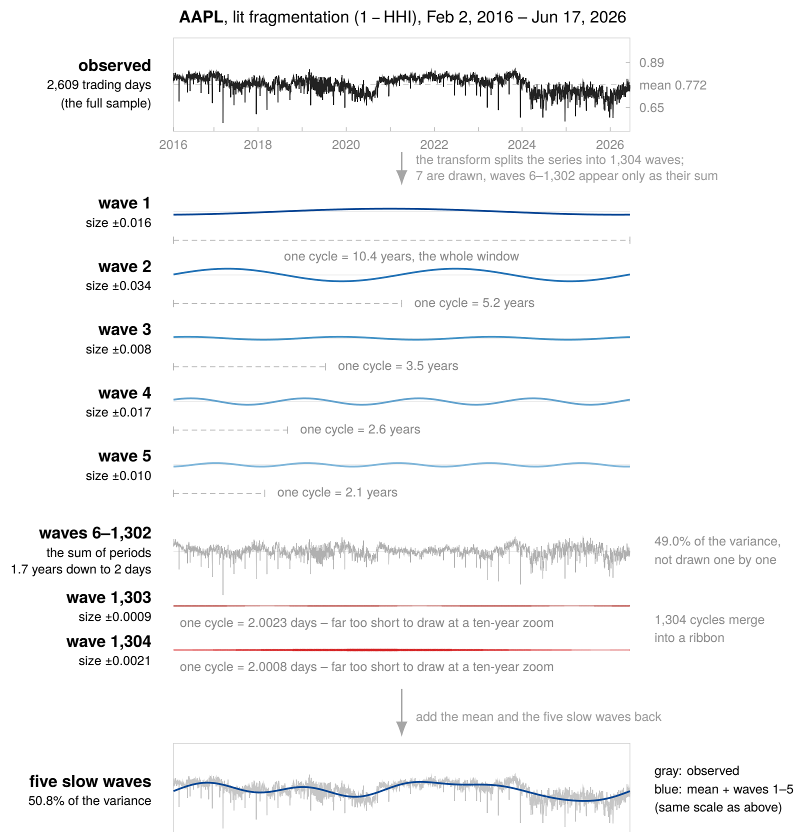
Panel A				
	Median	Mean	Q_1	Q_3
Detrended, $k = 1$	+0.575	+0.560	+0.485	+0.634
Detrended, $k = 2$	+0.546	+0.531	+0.452	+0.612
Tapered, $p = 3$	+0.505	+0.507	+0.365	+0.645
Logit, detrended $k = 1$	+0.585	+0.570	+0.497	+0.643
Logit, detrended $k = 2$	+0.559	+0.542	+0.463	+0.623
Logit, tapered $p = 3$	+0.517	+0.516	+0.374	+0.652
Panel B				
	Long memory, no breaks *simulated*	Venue fragmentation *data*	Pure level shifts *simulated*	Position (%)
Qu (2011) W, rejection at 5%	2.7	43.5	91.0	—
Panel C				
Full − First 2yr	+0.002	+0.030	+0.486	6
Full − Last 2yr	+0.009	+0.150	+0.465	31
Aggregated Difference	+0.017	+0.082	+0.258	27
Panel D				
$\hat{d}$, full series	+0.590	+0.590	—	—
Level shifts detected per series	9.47	7.17	—	—
$\hat{d}$, most significant shift removed	+0.509	+0.544	—	—
$\hat{d}$, all detected shifts removed	+0.098	+0.121	—	—

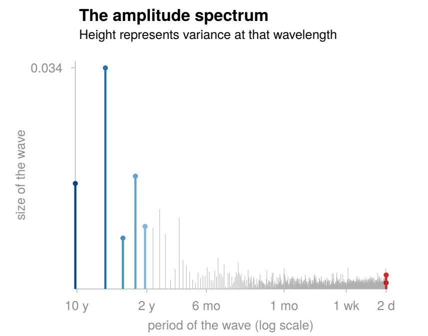
B The Fourier Decomposition of Fragmentation



The Fourier transformation can take any time series n and decompose it into $n/2$ waves and retain all information. Each wave carries a size and an alignment (its phase, drawn but not printed), except the fastest 2-day wave, the Nyquist Wave, which carries only a size. If you add all the waves and add the mean the observed series returns. In the paper, I set certain waves to zero, then transform the data, effectively "muting" that frequency. The Data used here is the daily fragmentation (1 – HHI) for AAPL, August 4 to November 2, 2020.

Figure A1: The transform of a 64-day window, wave by wave.





The slowest five waves are in blue and compose 50.8% of the variance. The red stems are the two fastest waves. The gray waves are the 1,297 waves between which contain 49.0% of the variance. At this width the 1,304 stems merge into a silhouette.

The Fourier transformation can take any time series n and decompose it into $n/2$ waves and retain all information. Each wave carries a size and an alignment (its phase, drawn but not printed). Here n is odd, so there is no 2-day Nyquist wave and all 1,304 waves carry both: 1 + 1,304 × 2 = 2,609 numbers out for 2,609 observations in. If you add all the waves and add the mean the observed series returns. In the paper, I set certain waves to zero, then transform the data, effectively "muting" that frequency. The Data used here is the daily fragmentation (1 – HHI) for AAPL, Feb 2, 2016 – Jun 17, 2026, the paper's full canonical window.

Figure A2: The same transform on the full 2,609-day sample.

Appendix Figures

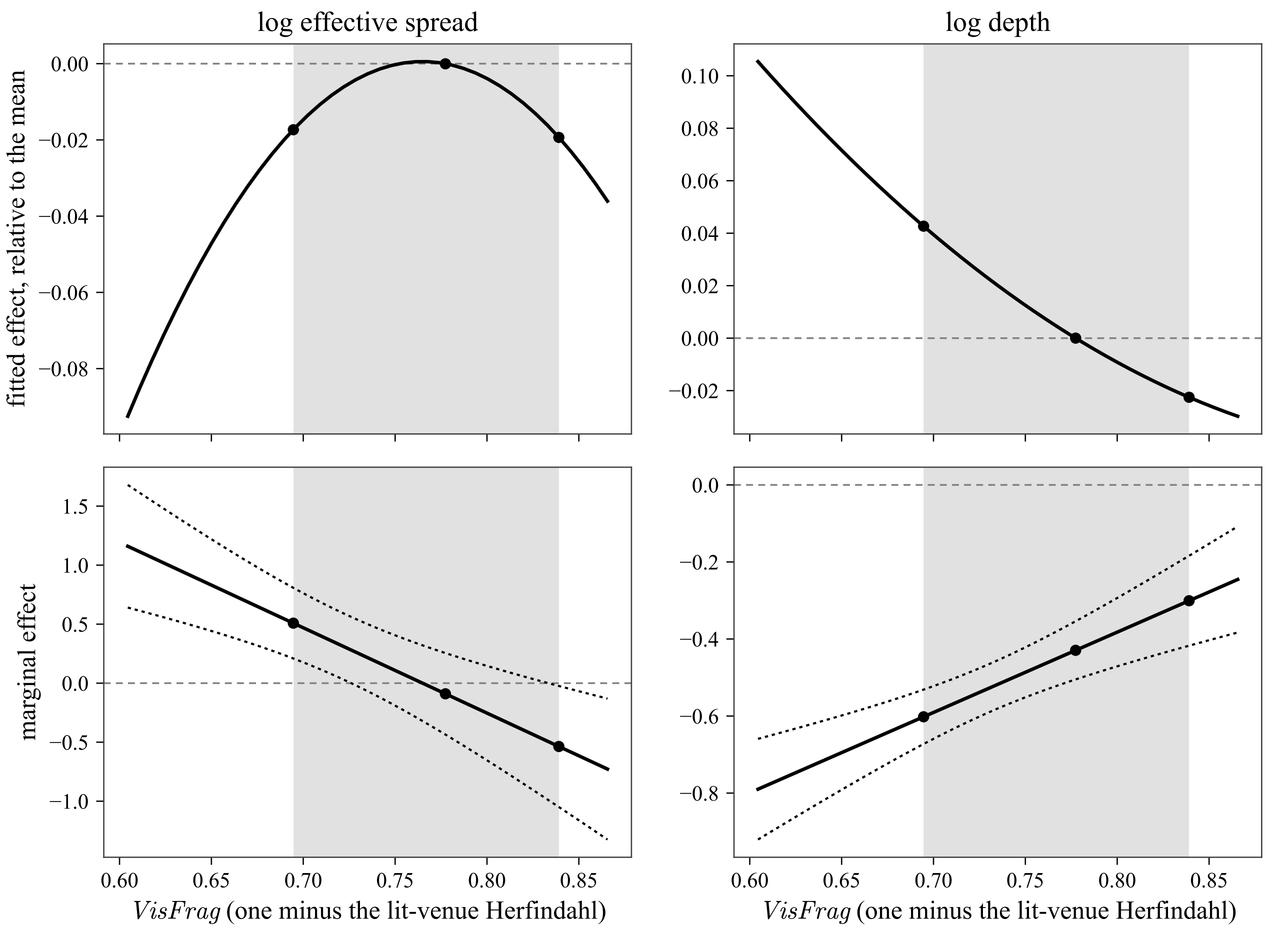
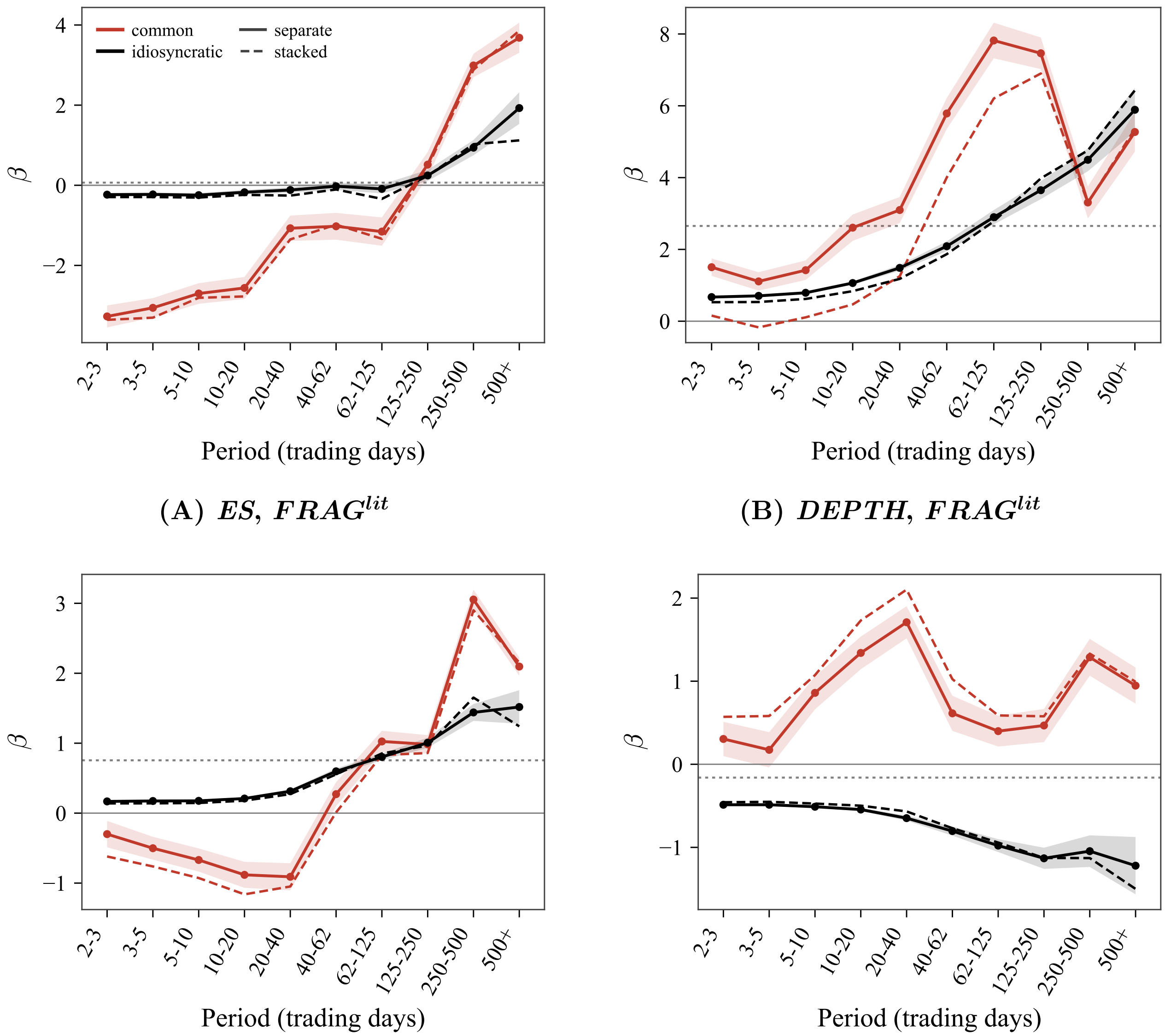


Figure A3: **The Estimated Effect of Fragmentation Across Its Range.** Degryse, De Jong and Kervel (2015) (DDK) includes a squared term $(FRAG^{lit})^2$ and instruments the three endogenous variables $FRAG^{lit}$, $(FRAG^{lit})^2$ and $dark$ using three instruments which are leave-one-out quartile size group averages of $FRAG^{lit}$, $(FRAG^{lit})^2$ and $dark$. The system of equations can be represented as (1) $FRAG^{lit}_{i,t} = a_{1,i}\delta_{q(t)} + b_1' X_{i,t} + c_1' W_{i,t} + \epsilon_{1,i,t}$, (2) $(FRAG^{lit}_{i,t})^2 = a_{2,i}\delta_{q(t)} + b_2' X_{i,t} + c_2' W_{i,t} + \epsilon_{2,i,t}$, (3) $dark_{i,t} = a_{3,i}\delta_{q(t)} + b_3' X_{i,t} + c_3' W_{i,t} + \epsilon_{3,i,t}$. The regression then estimates $Y_{i,t} = \alpha_i \delta_{q(t)} + \beta_1 \widehat{FRAG^{lit}}_{i,t} + \beta_2 (\widehat{FRAG^{lit}})^2_{i,t} + \beta_3 \widehat{dark}_{i,t} + \gamma' W_{i,t} + e_{i,t}$. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as $AvgLiq$; market cap, in log units, denoted as $Size$; the inverse price, in log units, denoted as $InversePrice$; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable $Algo$ in DDK's specification is standardized message traffic, as I do not have message data I proxy with NBBO updates divided by trading volume but note the limitation. $\alpha_i \delta_{q(t)}$ are stock-quarter fixed effects. DDK use Newey-West standard errors, but I use stock clustered. All variables in the DDK system are fractionally differenced at each stock's estimate of $\hat{d}^{FRAG}$, after differencing the first 250 observations of each stock's series are removed as they do not have significant pre-differencing observations to account for the long memory. The fractional differencing is applied before the stock-quarter demeaning. The system of equations is solved using GMM. Full tables for the DDK specification without differencing, without the quadratic specification and with controls not truncated are in the online appendix Table A6, and Table A7. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days. Because that specification enters fragmentation quadratically the estimated effect is not a number but a curve. The top row of panels are the fitted effect on liquidity, $\hat{\beta}_1(x - \bar{x}) + \hat{\beta}_2(x^2 - \bar{x}^2)$, measured against the sample mean of $FRAG^{lit}$. The bottom row panels are the slopes, the marginal effect, of which Table 7 reports three points. Dotted lines in the bottom row are 95% intervals computed using the delta method. Points mark the 10th percentile, the sample mean and the 90th percentile of $FRAG^{lit}$; the shaded region spans the 10th to 90th percentiles.



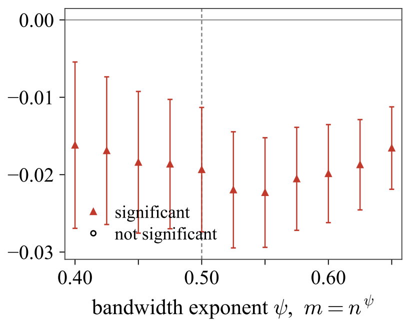
(A) *ES*, *FRAG*^{lit}

(B) *DEPTH*, *FRAG*^{lit}

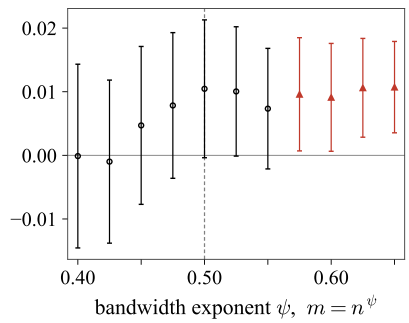
(C) *ES*, *dark*

(D) *DEPTH*, *dark*

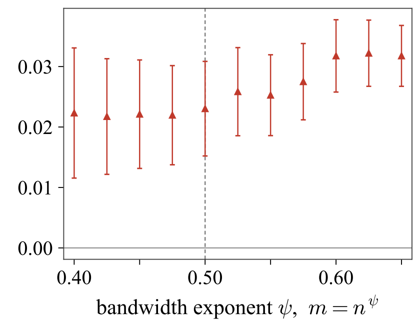
Figure A4: **Relation of Liquidity and Fragmentation Across Periods Decomposed by Components, Estimated as One Stacked Regression.** Each panel presents the coefficients from a panel regression of a liquidity measure, ES or $DEPTH$, on a fragmentation measure, $FRAG^{lit}$or $dark$, with control variables, decomposed between idiosyncratic and common component relations. Each regression is estimated on a panel that is fitted to the values of a pooled group of periods for one of the components, either common or idiosyncratic. The full ranges of the ten bands are [2-3), [3-5), [5-10), [10-20), [20-40), [40-62), [62-125), [125-250), [250-500), [500, 2,609) days. The regression estimated is $Y_{i,t}^{\mathcal{B},s} = \beta_1^{\mathcal{B},s} FRAG_{i,t}^{lit,\mathcal{B},s} + \beta_2^{\mathcal{B},s} dark_{i,t}^{\mathcal{B},s} + \gamma^{\mathcal{B},s\prime} Z_{i,t}^{\mathcal{B},s} + \varepsilon_{i,t}^{\mathcal{B},s}, \qquad s \in \{c, \iota\}$ where $\mathcal{B} \in \{1, 2, \ldots, 10\}$ and $Z^{\mathcal{B},s}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. $FRAG^{lit}$is 1 minus the HHI of trading volume shares using only lit exchanges and *dark* is the share of off-exchange volume. To compute the fitted panels which the regression is run on, first each stock is demeaned, then let z represent any variables time series. z is then decomposed into its idiosyncratic and common components as follows $\ddot{z}_{i,t} = a_i^z + \phi_i^{z\prime}\bar{h}_t + u_{i,t}^z, \qquad z_{i,t}^c = \hat{a}_i^z + \hat{\phi}_i^{z\prime}\bar{h}_t, \qquad z_{i,t}^\iota = \hat{u}_{i,t}^z$ where c denotes the common and ι denotes the idiosyncratic components. $\bar{h}_t$ is the vector of day t cross-sectional averages of every variable in the system: the liquidity measure, $FRAG^{lit}$, *dark*, the four control variables, and the leave-one-out size group averages of $FRAG^{lit}$, $FRAG^{lit}$squared and *dark*. Then a FFT is run on each stock's time series, this yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \ldots, \frac{n}{2}\}$ and n is the number of daily observations for stock i. The FFT is performed separately on the idiosyncratic and common component stock series. The periods are then pooled into the ten stated bands, separately for common and idiosyncratic components, and OLS is estimated on the ten fitted bands for each component. Each coefficient is from one of the forty estimations, ten bands for each component and each of the two liquidity measures. Standard errors are clustered on stock and date. The figure presents dotted lines to represent a regression which estimates $Y_{i,t}^{B,c+\iota}$ as the outcome and both the idiosyncratic and common component regressors entered jointly. The solid lines represents the coefficients from the main specification as in Figure 7 where the common and idiosyncratic components of the outcome and regressor panels are estimated separately "stacked". Red markers indicate coefficients for the common series and black markers indicate coefficients for the idiosyncratic series. The shaded region on the solid lines display the 95% confidence intervals. The sample is 3,630,896 stock days over 1,672 stocks during 2,609 trading days.



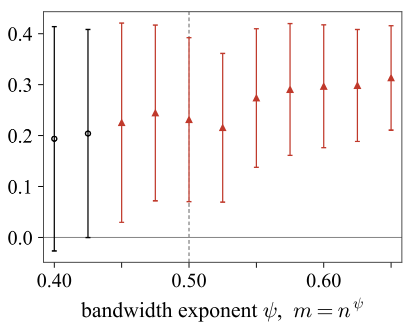
(A) log inverse price



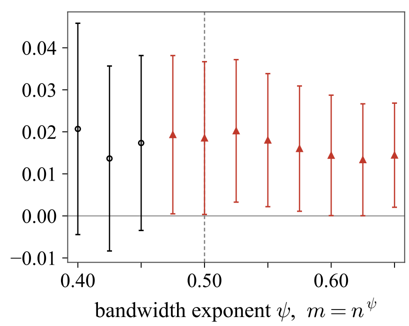
(B) log realized variance



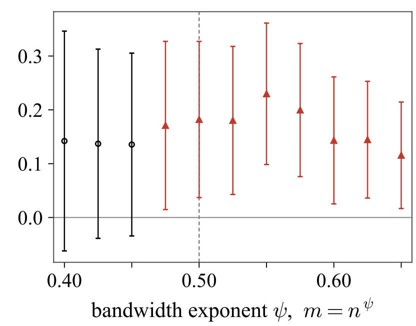
(C) log share volume



(D) off-exchange share

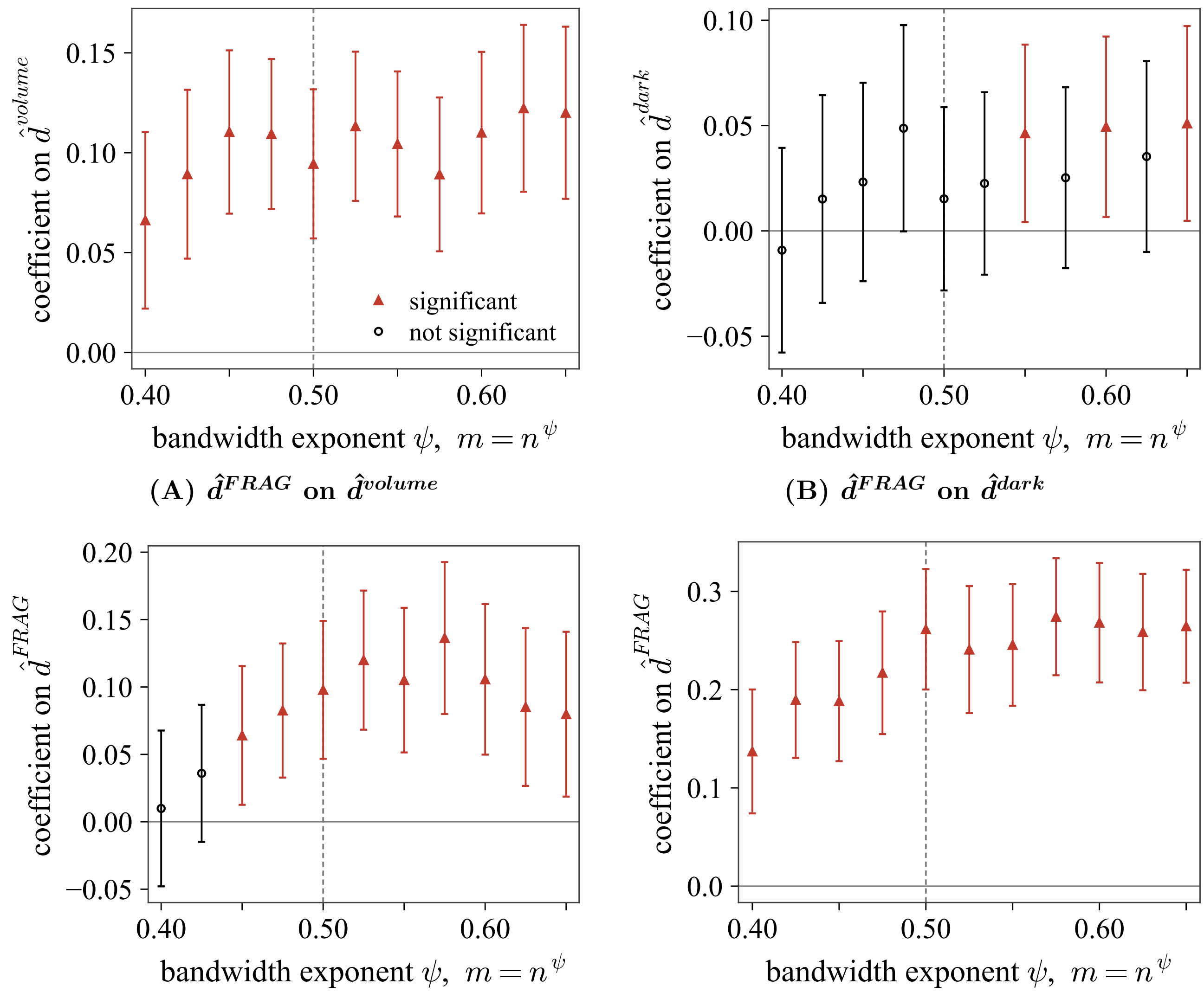


(E) tick-constrained (0/1)



(F) mean lit fragmentation (level)

Figure A5: **Bandwidth sensitivity of the cross-section of persistence.** Each panel plots the coefficient on one stock characteristic from the cross-sectional regression of $\hat{d}^{dark}$ on all 6 characteristics jointly. The x-axis is reestimations separately at differing bandwidths $m = n^{\psi}$, $\psi \in [0.40, 0.65]$ (m from 23 to 166). $\hat{d}^{dark}$ is the ELW (Shimotsu, 2010) estimate for *dark*. The dashed vertical rule marks the bandwidth $\psi = 1/2$ ($m = 51$) used throughout the paper. Whiskers are 95% confidence intervals from heteroskedasticity robust standard errors. Hollow black circles mark bandwidths at which the coefficient is not distinguishable from zero at the 5% level and solid red triangles mark those at which it is.



(A) $\hat{d}^{FRAG}$ on $\hat{d}^{volume}$

(B) $\hat{d}^{FRAG}$ on $\hat{d}^{dark}$

(C) $\hat{d}^{ES}$ on $\hat{d}^{FRAG}$

(D) $\hat{d}^{DEPTH}$ on $\hat{d}^{FRAG}$

Figure A6: **The association of memory parameters across bandwidths.** Each panel in the figure presents the coefficient from a regression of $\hat{d}^y$ on $\hat{d}^x$ with control variables. $FRAG^{lit}$ is 1 minus the HHI of volume shares using all lit exchanges and *dark* is the share of dark volume relative to all volume. There is one observation per stock. The estimates for $\hat{d}$ are from Shimotsu (2010) ELW. The control characteristics are stock means over the sample window. The characteristics are: *inverse price*, the inverse of price, in log units; *volatility*, computed as the sum of squared minute-level midpoint returns, in log units; *size*, measured as the shares outstanding multiplied by the median midpoint of that day, in log units; *volume*, the daily share volume, in log units; *dark*, the share of volume off-exchange; *tick constrained*, an indicator for tick constrained stocks, constructed following the definition of Kwan, Masulis and McInish (2015); and *fragmentation*, the raw fragmentation level. Whiskers are 95% confidence intervals from HC1 standard errors. Hollow black circles mark bandwidths at which the coefficient is not distinguishable from zero at the 5% level and solid red triangles mark those at which it is. The figure presents a bandwidth sensitivity. Each point on the x-axis is a coefficient measured with a different bandwidth m where $m = n^{\psi}$, n is the number of observations in the time series.

Appendix Tables

Table A2: Simulations of Long Memory With Trends and Level Shifts

The table presents Monte Carlo simulations for long memory and white noise time series. In panel A I simulate long memory using an ARFIMA$(0, 0.5, 0)$ model, over 200 simulated series of length $n = 2,609$. Series are generated by fractionally integrating Gaussian white noise under the Type II definition (Shimotsu and Phillips, 2005, eq. 2). The linear trend adds one standard deviation over the course of the simulated sample scaled linearly. The quadratic trends add one standard deviation at the center of the time series with a quadratic trend. The level shift adds one standard deviation at the midpoint of the series. The Raw column presents ELW estimates from Shimotsu (2010). The Detrended $k = 1$ and $k = 2$ rows present the ELW estimates on the residuals after detrending at order k. The tapered row follows Velasco (1999) for tapering the series and then estimating ELW. Panel B presents a series that has no long memory, the same detrending and tapering applications as in Panel A are applied.

	d_0	Raw	Detr. $k = 1$	Detr. $k = 2$	Taper $p = 3$
Panel A: Simulated Long Memory, $d_0 = 0.5$					
ARFIMA$(0, 0.5, 0)$, no contamination	0.50	0.511	0.483	0.457	0.493
+ linear trend	0.50	0.536	0.482	0.461	0.504
+ quadratic trend	0.50	0.515	0.468	0.445	0.463
+ one level shift	0.50	0.570	0.505	0.481	0.521
Panel B: No Simulated Memory, $d_0 = 0$					
White noise + linear trend	0.00	0.405	-0.036	-0.049	0.001
White noise + one level shift	0.00	0.567	0.408	0.407	0.549

Table A3: Cross-sectional Regressions of Persistence in *dark*

The table presents cross-sectional regressions of the stock-level memory parameter $\hat{d}^{dark}$ on stock characteristics using one observation per stock. $\hat{d}^{dark}$ is the Shimotsu (2010) two-step ELW estimate. Characteristics are stock means over the sample window. The characteristics are: *inverse price*, the inverse of price, in log units; *volatility*, computed as the sum of squared minute-level midpoint returns, in log units; *size*, measured as the shares outstanding multiplied by the median midpoint of that day, in log units; *volume*, the daily share volume, in log units; *dark*, the share of volume off-exchange; *tick constrained*, an indicator for tick constrained stocks, constructed following the definition of Kwan, Masulis and McInish (2015); and *fragmentation*, the raw fragmentation level. The estimates in columns (1)-(9) use $m = n^{1/2}$ and column (10) uses $m = n^{0.4}$. VIFs for the specification in (8) are displayed in (11) and VIFs for the specification in (9) are presented in (12). Heteroskedasticity robust t-statistics in parentheses; *, ** and *** denote significance at the 10%, 5% and 1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Univariate							*Full*	*Reduced*	*Reduced*	*VIF*	
								$n^{1/2}$	$n^{1/2}$	$n^{0.4}$	Full	Reduced
Inverse price	-0.0022							-0.0116*	-0.0210***	-0.0160***	7.8	2.9
	(-0.82)							(-1.69)	(-5.15)	(-2.92)		
Volatility		0.0072**						0.0169**	0.0103*	-0.0007	4.0	2.6
		(1.98)						(2.39)	(1.86)	(-0.10)		
Size			0.0121***					0.0110			12.2	
			(6.84)					(1.63)				
Volume				0.0337***				0.0130*	0.0226***	0.0220***	5.5	1.8
				(10.84)				(1.74)	(5.68)	(4.01)		
Dark share					0.4133***			0.2609***	0.2552***	0.2014*	1.9	1.9
					(6.70)			(3.19)	(3.11)	(1.80)		
Tick constrained						0.0295***		0.0204**	0.0213**	0.0213*	2.2	2.2
						(4.26)		(2.20)	(2.30)	(1.66)		
Fragmentation							0.3815***	0.2134***	0.2020***	0.1515	1.3	1.3
							(5.49)	(2.80)	(2.66)	(1.46)		
R^2	0.000	0.002	0.022	0.069	0.033	0.012	0.017	0.093	0.091	0.040		
N	1,753	1,753	1,753	1,753	1,753	1,753	1,753	1,753	1,753	1,725		

Table A4: The persistence of venue fragmentation and the persistence of liquidity

The table presents cross-sectional regressions of $\hat{d}^y$ on $\hat{d}^x$ with control variables. $FRAG^{lit}$is 1 minus the HHI of volume shares using all lit exchanges and *dark* is the share of dark volume relative to all volume. There is one observation per a stock. The estimates for $\hat{d}$ are from Shimotsu (2010) ELW. The control characteristics are stock means over the sample window. The characteristics are: *inverse price*, the inverse of price, in log units; *volatility*, computed as the sum of squared minute-level midpoint returns, in log units; *size*, measured as the shares outstanding multiplied by the median midpoint of that day, in log units; *volume*, the daily share volume, in log units; *dark*, the share of volume off-exchange; *tick constrained*, an indicator for tick constrained stocks, constructed following the definition of Kwan, Masulis and McInish (2015). *t*-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels.

	$\hat{d}^{FRAG}$				$\hat{d}^{ES}$		$\hat{d}^{DEPTH}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{d}^{volume}$	0.1393***	0.0961***						
	(7.35)	(5.02)						
$\hat{d}^{dark}$			0.0266	0.0169				
			(1.18)	(0.76)				
$\hat{d}^{FRAG}$					0.2661***	0.0979***	0.2055***	0.2633***
					(8.94)	(3.75)	(6.80)	(8.61)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.031	0.136	0.001	0.124	0.042	0.407	0.027	0.107
N	1,753	1,753	1,753	1,753	1,750	1,750	1,754	1,754

Table A5: Aggregate and stock-level memory of venue fragmentation

The aggregate is the equal-weighted cross-sectional mean of venue fragmentation on each day, transformed before averaging, and $\hat{d}$ is estimated on that single daily series. The stock median is the median of the stock-by-stock estimates over the same stocks, the same transform and the same trend treatment; comparing across transforms would not be like for like. Estimator and bandwidth are those of Table 2: two-step exact local Whittle at $m = n^{1/2}$, and the tapered rows use the order-3 taper of Velasco (1999). The gap is scaled by the *aggregate's* standard error, $1/(2\sqrt{m}) = 0.0700$ at $m = 51$. The two sides are not symmetric: the stock median is a median of roughly seventeen hundred estimates and is tightly determined, while the aggregate is a single estimate of a single series, so the uncertainty in the difference is essentially the aggregate's own. Panel B addresses composition. An equal-weighted average over a membership that changes from day to day can drift for compositional reasons alone, and coverage here runs from 1,283 to 1,617 stocks per day. Panel B repeats the exercise on the stocks present on every day of the window, where that channel is shut off by construction.

	Aggregate $\hat{d}$	Stock median $\hat{d}$	Gap	Gap / s.e.
Panel A: all stocks present each day				
Raw fragmentation				
No trend removal	0.7337	0.5816	0.1521	2.2
Detrended, $k = 1$	0.6885	0.5748	0.1137	1.6
Detrended, $k = 2$	0.6840	0.5461	0.1379	2.0
Tapered, $p = 3$	0.5912	0.5052	0.0860	1.2
Logit fragmentation				
No trend removal	0.7660	0.5906	0.1754	2.5
Detrended, $k = 1$	0.7053	0.5853	0.1199	1.7
Detrended, $k = 2$	0.7028	0.5592	0.1437	2.1
Tapered, $p = 3$	0.6193	0.5166	0.1027	1.5
Panel B: balanced panel, 745 stocks present every day				
Raw fragmentation				
No trend removal	0.6725	0.5784	0.0941	1.3
Detrended, $k = 1$	0.6561	0.5728	0.0833	1.2
Detrended, $k = 2$	0.6533	0.5540	0.0993	1.4
Tapered, $p = 3$	0.4884	0.4582	0.0302	0.4
Logit fragmentation				
No trend removal	0.7002	0.5875	0.1127	1.6
Detrended, $k = 1$	0.6664	0.5872	0.0792	1.1
Detrended, $k = 2$	0.6641	0.5747	0.0894	1.3
Tapered, $p = 3$	0.5076	0.4707	0.0369	0.5

Table A6: Estimations from the Degryse, De Jong and Kervel (2015) Specification, Full Coefficients

Degryse, De Jong and Kervel (2015) (DDK) includes a squared term $(FRAG^{lit})^2$ and instruments the three endogenous variables $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark* using three instruments which are leave-one-out quartile size group averages of $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark*. The system of equations can be represented as (1) $FRAG^{lit}_{i,t} = a_{1,i}\delta_{q(t)} + b_1'X_{i,t} + c_1'W_{i,t} + \epsilon_{1,i,t}$, (2) $(FRAG^{lit}_{i,t})^2 = a_{2,i}\delta_{q(t)} + b_2'X_{i,t} + c_2'W_{i,t} + \epsilon_{2,i,t}$, (3) $dark_{i,t} = a_{3,i}\delta_{q(t)} + b_3'X_{i,t} + c_3'W_{i,t} + \epsilon_{3,i,t}$. The regression then estimates $Y_{i,t} = \alpha_i\delta_{q(t)} + \beta_1\widehat{FRAG^{lit}}_{i,t} + \beta_2\widehat{(FRAG^{lit})^2}_{i,t} + \beta_3\widehat{dark}_{i,t} + \gamma'W_{i,t} + e_{i,t}$. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as *AvgLiq*; market cap, in log units, denoted as *Size*; the inverse price, in log units, denoted as $InversePrice$; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable *Algo* in DDK's specification is standardized message traffic, as I do not have message data I proxy with NBBO updates divided by trading volume but note the limitation. $\alpha_i\delta_{q(t)}$ are stock-quarter fixed effects. DDK use Newey-West standard errors, but I use stock clustered. I report the marginal effect at the sample mean as well as the 10^{th} and 90^{th} percentiles. All variables in the DDK system are fractionally differenced at each stock's estimate of $\hat{d}^{FRAG}$, after differencing the first 250 observations of each stock's series are removed as they do not have significant pre-differencing observations to account for the long memory. The fractional differencing is applied before the stock-quarter demeaning. The system of equations is solved using GMM. Full tables for the DDK specification without differencing, without the quadratic specification and with controls not truncated are in the online appendix Table A6, and Table A7. *t*-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days. This table reports the same estimations as Table 7 with the control coefficients no longer truncated. The *levels* columns estimate the system on the raw series and the $\Delta^{\hat{d}}$ columns on the fractionally differenced series.

	ES		*DEPTH*	
	levels	$\Delta^{\hat{d}}$	levels	$\Delta^{\hat{d}}$
VisFrag	9.5838***	5.5305***	-0.4271	-2.0504***
	(7.61)	(4.04)	(-0.76)	(-6.02)
VisFrag^2	-6.2352***	-3.6151***	-0.0950	1.0426***
	(-7.45)	(-3.84)	(-0.25)	(4.49)
Dark	-1.3521***	-0.7503***	0.8912***	0.8035***
	(-37.01)	(-18.32)	(18.78)	(28.20)
$AvgLiq_{-i,t}$	0.8416***	0.8999***	0.9422***	1.0183***
	(92.73)	(73.35)	(93.84)	(114.78)
Ln Size	-0.0785***	-0.0074	0.3363***	0.1701***
	(-3.42)	(-0.42)	(6.80)	(6.00)
Inverse price	0.2646***	0.3099***	0.1287**	-0.0001
	(10.34)	(15.35)	(2.35)	(-0.00)
Ln Volume	-0.0627***	-0.0341***	0.1222***	0.1079***
	(-38.81)	(-19.45)	(65.70)	(72.59)
Ln SD	0.0214***	0.0606***	-0.0229***	-0.0062***
	(16.10)	(32.11)	(-8.98)	(-2.80)
Algo	-0.1957***	-0.1333***	0.0798***	0.1045***
	(-72.13)	(-45.10)	(17.50)	(31.17)

Table A6: Estimations from the Degryse, De Jong and Kervel (2015) Specification, Full Coefficients (continued)

	ES		*DEPTH*	
	levels	$\Delta^{\hat{d}}$	levels	$\Delta^{\hat{d}}$
Marginal effect of fragmentation				
at the 10th percentile	0.9217***	0.5084***	-0.5591***	-0.6019***
	(6.31)	(3.33)	(-9.96)	(-16.72)
at the sample mean	-0.1103	-0.0900	-0.5748***	-0.4294***
	(-0.89)	(-0.51)	(-10.83)	(-11.14)
at the 90th percentile	-0.8803***	-0.5364**	-0.5866***	-0.3006***
	(-4.66)	(-2.05)	(-6.71)	(-5.05)
Fixed effects	stock-quarter			
First-stage F, *VisFrag*	45.5	41.7	46.4	41.8
First-stage F, $VisFrag^2$	63.7	56.3	64.9	55.5
First-stage F, *Dark*	12,868.5	10,995.7	13,698.0	11,613.6
R^2	0.838	0.060	0.940	0.814
N	3,773,606	3,334,106	3,773,606	3,334,106
Stocks	1,758			

Table A7: Estimations from the Degryse, De Jong and Kervel (2015) Specification without the Quadratic Term, Full Coefficients

Degryse, De Jong and Kervel (2015) (DDK) includes a squared term $(FRAG^{lit})^2$ and instruments the three endogenous variables $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark* using three instruments which are leave-one-out quartile size group averages of $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark*. The system of equations can be represented as (1) $FRAG^{lit}_{i,t} = a_{1,i}\delta_{q(t)} + b_1' X_{i,t} + c_1' W_{i,t} + \epsilon_{1,i,t}$, (2) $(FRAG^{lit}_{i,t})^2 = a_{2,i}\delta_{q(t)} + b_2' X_{i,t} + c_2' W_{i,t} + \epsilon_{2,i,t}$, (3) $dark_{i,t} = a_{3,i}\delta_{q(t)} + b_3' X_{i,t} + c_3' W_{i,t} + \epsilon_{3,i,t}$. The regression then estimates $Y_{i,t} = \alpha_i \delta_{q(t)} + \beta_1 \widehat{FRAG^{lit}}_{i,t} + \beta_2 (\widehat{FRAG^{lit}})^2_{i,t} + \beta_3 \widehat{dark}_{i,t} + \gamma' W_{i,t} + e_{i,t}$. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as $AvgLiq$; market cap, in log units, denoted as $Size$; the inverse price, in log units, denoted as $InversePrice$; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable $Algo$ in DDK's specification is standardized message traffic, as I do not have message data I proxy with NBBO updates divided by trading volume but note the limitation. $\alpha_i \delta_{q(t)}$ are stock-quarter fixed effects. DDK use Newey-West standard errors, but I use stock clustered. I report the marginal effect at the sample mean as well as the 10^{th} and 90^{th} percentiles. All variables in the DDK system are fractionally differenced at each stock's estimate of $\hat{d}^{FRAG}$, after differencing the first 250 observations of each stock's series are removed as they do not have significant pre-differencing observations to account for the long memory. The fractional differencing is applied before the stock-quarter demeaning. The system of equations is solved using GMM. Full tables for the DDK specification without differencing, without the quadratic specification and with controls not truncated are in the online appendix Table A6, and Table A7. *t*-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days. This table reports the same estimations as Table A6 with $(FRAG^{lit})^2$ dropped as a regressor and its leave-one-out group average retained as an instrument, the robustness specification of Degryse, De Jong and Kervel (2015, footnote 22, p. 1611). With the quadratic removed the coefficient on $FRAG^{lit}$ is the marginal effect itself, constant by construction, so no marginal effects are reported.

	ES		*DEPTH*	
	levels	$\Delta^{\hat{d}}$	levels	$\Delta^{\hat{d}}$
VisFrag	0.2872***	0.3173**	-0.5677***	-0.5307***
	(2.74)	(2.32)	(-12.67)	(-17.06)
Dark	-1.2795***	-0.7368***	0.8961***	0.7828***
	(-36.63)	(-18.27)	(20.76)	(27.81)
$AvgLiq_{-i,t}$	0.8414***	0.9007***	0.9423***	1.0163***
	(94.76)	(74.66)	(93.87)	(114.53)
Ln Size	-0.0877***	0.0024	0.3363***	0.1371***
	(-3.90)	(0.14)	(6.80)	(5.00)
Inverse price	0.2551***	0.3184***	0.1289**	-0.0346
	(10.17)	(15.89)	(2.36)	(-0.96)
Ln Volume	-0.0632***	-0.0348***	0.1222***	0.1076***
	(-39.93)	(-20.30)	(66.55)	(72.21)
Ln SD	0.0225***	0.0611***	-0.0229***	-0.0058***
	(17.36)	(32.78)	(-8.98)	(-2.61)
Algo	-0.1908***	-0.1321***	0.0803***	0.1022***
	(-74.29)	(-45.72)	(19.41)	(30.87)
Fixed effects	stock-quarter			
First-stage F, *VisFrag*	4,173.0	2,487.7	3,956.6	2,356.9
First-stage F, *Dark*	6,537.9	5,498.1	6,904.3	5,811.9
R^2	0.842	0.068	0.940	0.814
N	3,773,606	3,334,106	3,773,606	3,334,106
Stocks	1,758			

Table A8: Estimations from the Ergemen and Velasco (2017) Specification, Full Coefficients

The table presents the coefficient estimates from regressions of liquidity variables *ES* and *DEPTH* on fragmentation variables $FRAG^{lit}$and *dark*. *ES* is the volume weighted daily effective spread computed using the Battalio et al. (2026) LF-NBBO and signing methodology, and *DEPTH* is time weighted sum of the depth at the best LF bid and ask price. Both liquidity outcome variables are in log units. $FRAG^{lit}$is 1 minus the HHI of exchange venue volume shares, using only lit exchanges. *dark* is the volume share of off-exchange volume. The control variables used in Ergemen and Velasco (2017) (EV) are: the daily share volume, denoted *volume*, in log units; the inverse price, denoted *price* in log units; the daily volatility, computed as the rolling 20-day standard deviation of log returns, denoted *volatility*, in log units; and the market capitalization, denoted *size* in log units. The steps of EV are as follows 1. all variables in the system are first differenced 2. all first differenced variables in the system are regressed on an intercept and a vector $\bar{H}_t$ which contains the day t cross-sectional averages of all variables *at the stock level* so all stocks are allowed heterogeneous loadings on the common factors 3. the residuals from the stock level regressions are pooled and the coefficients are estimated. The specification can be modeled by $\Delta Y_{i,t} = a_i + \beta_1 \Delta FRAG^{lit}_{i,t} + \beta_2 \Delta dark_{i,t} + \gamma' \Delta Z_{i,t} + \phi'_i \bar{H}_t + \varepsilon_{i,t}$. Standard errors are clustered two ways, on stock and on date (Cameron, Gelbach and Miller, 2011). The projection removes the average factor loading and not its dispersion, so the scores still co-move within a day; clustering on stock alone understates the standard error by 6% to 26% on these coefficients. Table A8 presents the full EV specification without controls truncated. This table reports the same estimation as Table 7 with the control coefficients no longer truncated.

	ES	*DEPTH*
$FRAG_{\text{lit}}$	-0.2390***	0.5629***
	(-10.87)	(55.35)
dark	0.1355***	-0.4701***
	(17.45)	(-99.50)
Log volume	0.0037*	0.1155***
	(1.65)	(60.16)
Inverse price	0.3874***	0.0084
	(18.63)	(0.25)
Log return volatility	0.1522***	-0.0513***
	(35.00)	(-16.33)
Log market capitalization	0.0107	0.1296***
	(0.68)	(5.29)
Estimator	pooled CCP, differenced and defactored	
R^2	0.076	0.141
N	3,757,425	3,757,425
Stocks	1,758	

Table A9: Altering the Instruments in the Degryse, De Jong and Kervel (2015) Specification

Degryse, De Jong and Kervel (2015) (DDK) includes a squared term $(FRAG^{lit})^2$ and instruments the three endogenous variables $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark* using three instruments which are leave-one-out quartile size group averages of $FRAG^{lit}$, $(FRAG^{lit})^2$ and *dark*. The system of equations can be represented as (1) $FRAG^{lit}_{i,t} = a_{1,i}\delta_{q(t)} + b_1' X_{i,t} + c_1' W_{i,t} + \epsilon_{1,i,t}$, (2) $(FRAG^{lit}_{i,t})^2 = a_{2,i}\delta_{q(t)} + b_2' X_{i,t} + c_2' W_{i,t} + \epsilon_{2,i,t}$, (3) $dark_{i,t} = a_{3,i}\delta_{q(t)} + b_3' X_{i,t} + c_3' W_{i,t} + \epsilon_{3,i,t}$. The regression then estimates $Y_{i,t} = \alpha_i\delta_{q(t)} + \beta_1 \widehat{FRAG^{lit}}_{i,t} + \beta_2 (\widehat{FRAG^{lit}})^2_{i,t} + \beta_3 \widehat{dark}_{i,t} + \gamma' W_{i,t} + e_{i,t}$. $W_{i,t}$ is DDK's control set including: leave-one-out average liquidity of the same size group (ES or $DEPTH$) denoted as $AvgLiq$; market cap, in log units, denoted as $Size$; the inverse price, in log units, denoted as $InversePrice$; share volume, in log units, denoted as $Volume$; the 20-day rolling standard deviation of log returns, in log units, denoted as $Volatility$; the final control variable $Algo$ in DDK's specification is standardized message traffic, as I do not have message data I proxy with NBBO updates divided by trading volume but note the limitation. $\alpha_i\delta_{q(t)}$ are stock-quarter fixed effects. DDK use Newey-West standard errors, but I use stock clustered. I report the marginal effect at the sample mean as well as the 10^{th} and 90^{th} percentiles. All variables in the DDK system are fractionally differenced at each stock's estimate of $\hat{d}^{FRAG}$, after differencing the first 250 observations of each stock's series are removed as they do not have significant pre-differencing observations to account for the long memory. The fractional differencing is applied before the stock-quarter demeaning. The system of equations is solved using GMM. Full tables for the DDK specification without differencing, without the quadratic specification and with controls not truncated are in the online appendix Table A6, and Table A7. t-statistics are in parentheses beneath each coefficient; *, ** and *** denote significance at the 10%, 5% and 1% levels. The sample is 3,773,606 stock days over 1,758 stocks during 2,609 trading days. This table holds that specification fixed, the panel, the $\Delta^{\hat{d}}$ transform, the control set and the stock-clustered standard errors are those of Table 7 throughout, and varies only which variables are treated as endogenous and which leave-one-out group averages are used as excluded instruments. Each row is one such specification, so reading down a column shows which part of the instrument set the sign of the fragmentation coefficient depends on.

		log depth		log eff. spread	
		$FRAG_{\text{lit}}$ (1)	*dark* (2)	$FRAG_{\text{lit}}$ (3)	*dark* (4)
Panel A: $FRAG_{\text{lit}}$ only					
(1)	OLS	0.5191***		-0.3562***	
		(145.91)		(-38.75)	
(2)	2SLS, IV: lit	0.1823***		-0.2151*	
		(15.42)		(-1.67)	
(3)	2SLS, IV: lit lit^2	0.1826***		-0.2297*	
		(15.45)		(-1.78)	
(4)	2SLS, IV: dark	0.8562***		-1.1773***	
		(34.23)		(-8.94)	
(5)	2SLS, IV: lit dark	0.2260***		-0.4090***	
		(19.37)		(-3.19)	
(6)	2SLS, IV: lit lit^2 dark	0.2257***		-0.4384***	
		(19.35)		(-3.43)	
Panel B: dark only					
(7)	OLS		-0.3724***		0.0312***
			(-212.48)		(9.61)
(8)	2SLS, IV: dark		0.4877***		-0.5670***
			(34.95)		(-8.76)
(9)	2SLS, IV: lit		0.1998***		-0.2826*
			(15.09)		(-1.66)
(10)	2SLS, IV: dark lit		0.3545***		-0.6986***
			(30.90)		(-18.34)
(11)	2SLS, IV: dark lit lit^2		0.3526***		-0.7045***
			(30.82)		(-18.49)
Panel C: $FRAG_{\text{lit}}$ and dark					
(12)	OLS	0.6123***	-0.4098***	-0.3660***	0.0501***
		(169.12)	(-230.64)	(-40.20)	(15.97)
(13)	2SLS, endog: lit, IV: lit	0.5529***	-0.4062***	-0.2486*	0.0440***
		(43.76)	(-105.22)	(-1.89)	(4.78)
(14)	2SLS, endog: lit, IV: lit lit^2	0.5535***	-0.4057***	-0.2651**	0.0444***
		(43.81)	(-105.14)	(-2.02)	(4.82)
(15)	2SLS, endog: dark, IV: dark	0.4689***	0.2206***	-0.2709***	-0.4366***
		(53.16)	(14.02)	(-10.57)	(-6.81)
(16)	2SLS, endog: dark, IV: dark lit	0.4425***	0.0904***	-0.2755***	-0.6737***
		(50.50)	(6.08)	(-10.75)	(-17.44)
(17)	2SLS, endog: lit dark, IV: lit dark	-0.5472***	0.7995***	0.3417**	-0.7317***
		(-17.46)	(28.16)	(2.49)	(-18.12)
(18)	2SLS, endog: lit dark, IV: lit lit^2 dark	-0.5307***	0.7828***	0.3173**	-0.7368***
		(-17.06)	(27.81)	(2.32)	(-18.27)

Table A10: Relation of Liquidity and Fragmentation Across Periods

Each panel presents the coefficients from a panel regression of a liquidity measure, *ES* or *DEPTH*, on a fragmentation measure, $FRAG^{lit}$or *dark*, with control variables. Each regression is estimated on a panel that is fitted to the values of a pooled group of periods. The full ranges of the ten bands are [2-3), [3-5), [5-10), [10-20), [20-40), [40-62), [62-125), [125-250), [250-500), [500, 2,609) days. The regression estimated is $Y_{i,t}^{\mathcal{B}} = \beta_1^{\mathcal{B}} FRAG_{i,t}^{lit,\mathcal{B}} + \beta_2^{\mathcal{B}} dark_{i,t}^{\mathcal{B}} + \gamma^{\mathcal{B}\prime} Z_{i,t}^{\mathcal{B}} + \varepsilon_{i,t}^{\mathcal{B}}$ where $\mathcal{B} \in \{1, 2, \ldots, 10\}$ and $Z^{\mathcal{B}}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. $FRAG^{lit}$is 1 minus the HHI of trading volume shares using only lit exchanges and *dark* is the share of off-exchange volume. To compute the fitted panels which the regression is run on, first a FFT is run on each stock's time series, this yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \ldots, \frac{n}{2}\}$ and n is the number of daily observations for stock i. The periods are then pooled into the ten stated bands and OLS is estimated on the ten fitted bands. Each coefficient is from one of the twenty estimations, ten bands for each of the two liquidity measures. Standard errors are clustered on stock and date (Cameron, Gelbach and Miller, 2011). The FFT is mean zero by construction and takes the place of stock fixed effects. The dotted line presents the pooled coefficient, estimated on the same panel with the same control variables and no band-pass applied. Red triangles indicate significance at the 5% level and black circles are insignificant by the same standard. The whiskers display the 95% confidence intervals. The sample is 3,630,896 stock days over 1,672 stocks during 2,609 trading days. This table reports the coefficients plotted in Figure 6 and their t-statistics; *, ** and *** denote significance at the 10%, 5% and 1% levels.

	$FRAG^{lit}$		$dark$	
Period (days)	β	t	β	t
Panel A: ES				
2–3	−0.727***	−16.9	+0.025**	+2.1
3–5	−0.759***	−18.5	+0.026**	+2.4
5–10	−0.654***	−18.8	+0.017*	+1.8
10–20	−0.556***	−16.8	−0.033**	−2.5
20–40	−0.510***	−12.4	−0.013	−0.7
40–62	−0.355***	−6.6	+0.212***	+8.4
62–125	−0.614***	−8.3	+0.612***	+16.0
125–250	+0.417***	+4.7	+0.814***	+18.9
250–500	+1.597***	+15.0	+2.300***	+35.9
500+	+3.252***	+18.9	+1.995***	+31.6
Panel B: DEPTH				
2–3	+0.604***	+49.7	−0.401***	−72.0
3–5	+0.566***	+43.0	−0.397***	−69.3
5–10	+0.712***	+46.6	−0.361***	−51.7
10–20	+0.970***	+39.0	−0.267***	−21.8
20–40	+1.389***	+32.9	−0.180***	−8.8
40–62	+2.394***	+31.9	−0.249***	−7.5
62–125	+3.514***	+29.9	−0.354***	−8.0
125–250	+4.723***	+32.3	−0.406***	−5.5
250–500	+4.433***	+25.9	−0.021	−0.2
500+	+5.389***	+20.9	+0.749***	+7.1

Table A11: Relation of Liquidity and Fragmentation Across Periods Decomposed by Components

Each panel presents the coefficients from a panel regression of a liquidity measure, *ES* or *DEPTH*, on a fragmentation measure, $FRAG^{lit}$or *dark*, with control variables, decomposed between idiosyncratic and common component relations. Each regression is estimated on a panel that is fitted to the values of a pooled group of periods for one of the components, either common or idiosyncratic. The full ranges of the ten bands are [2-3), [3-5), [5-10), [10-20), [20-40), [40-62), [62-125), [125-250), [250-500), [500, 2,609) days. The regression estimated is $Y_{i,t}^{\mathcal{B},s} = \beta_1^{\mathcal{B},s} FRAG_{i,t}^{lit,\mathcal{B},s} + \beta_2^{\mathcal{B},s} dark_{i,t}^{\mathcal{B},s} + \gamma^{\mathcal{B},s\prime} Z_{i,t}^{\mathcal{B},s} + \varepsilon_{i,t}^{\mathcal{B},s}, \qquad s \in \{c, \iota\}$ where $\mathcal{B} \in \{1, 2, \ldots, 10\}$ and $Z^{\mathcal{B},s}$ is the same vector of controls that includes: the market capitalization, in log units, denoted *size*; the inverse price, in log units, denoted as *inverse price*; the rolling 20-day standard deviation of log returns, in log units, denoted as *volatility*; and share volume, in log units, denoted as *volume*. $FRAG^{lit}$is 1 minus the HHI of trading volume shares using only lit exchanges and *dark* is the share of off-exchange volume. To compute the fitted panels which the regression is run on, first each stock is demeaned, then let z represent any variables time series. z is then decomposed into its idiosyncratic and common components as follows $\ddot{z}_{i,t} = a_i^z + \phi_i^{z\prime} \bar{h}_t + u_{i,t}^z, \qquad z_{i,t}^c = \hat{a}_i^z + \hat{\phi}_i^{z\prime} \bar{h}_t, \qquad z_{i,t}^\iota = \hat{u}_{i,t}^z$ where c denotes the common and ι denotes the idiosyncratic components. $\bar{h}_t$ is the vector of day t cross-sectional averages of every variable in the system: the liquidity measure, $FRAG^{lit}$, *dark*, the four control variables, and the leave-one-out size group averages of $FRAG^{lit}$, $FRAG^{lit}$squared and *dark*. Then a FFT is run on each stock's time series, this yields j cycles of period length $\frac{n}{j}$ where $j \in \{1, 2, \ldots, \frac{n}{2}\}$ and n is the number of daily observations for stock i. The FFT is performed separately on the idiosyncratic and common component stock series. The periods are then pooled into the ten stated bands, separately for common and idiosyncratic components, and OLS is estimated on the ten fitted bands for each component. Each coefficient is from one of the forty estimations, ten bands for each component and each of the two liquidity measures. Standard errors are clustered on stock and date. The FFT is mean zero by construction and takes the place of stock fixed effects. The sample is 3,630,896 stock days over 1,672 stocks during 2,609 trading days. This table reports the coefficients plotted in Figure 7 and their t-statistics; *, ** and *** denote significance at the 10%, 5% and 1% levels.

	common part		idiosyncratic part		
Period (days)	β	(s.e.)	β	(s.e.)	difference z
Panel A: ES, $FRAG^{lit}$					
2–3	−3.273***	(0.140)	−0.234***	(0.024)	−22.3
3–5	−3.058***	(0.124)	−0.228***	(0.025)	−23.1
5–10	−2.701***	(0.131)	−0.246***	(0.028)	−18.9
10–20	−2.562***	(0.141)	−0.173***	(0.030)	−17.1
20–40	−1.072***	(0.161)	−0.117***	(0.035)	−6.1
40–62	−1.023***	(0.171)	−0.024	(0.047)	−6.0
62–125	−1.155***	(0.180)	−0.090	(0.058)	−5.9
125–250	+0.516***	(0.160)	+0.245***	(0.071)	+1.7
250–500	+2.990***	(0.149)	+0.939***	(0.097)	+13.2
500+	+3.682***	(0.192)	+1.927***	(0.201)	+7.9
Panel B: ES, dark					
2–3	−0.300***	(0.096)	+0.168***	(0.009)	−4.9
3–5	−0.502***	(0.083)	+0.174***	(0.009)	−8.1
5–10	−0.670***	(0.085)	+0.176***	(0.009)	−10.0
10–20	−0.882***	(0.095)	+0.209***	(0.010)	−11.5
20–40	−0.910***	(0.099)	+0.313***	(0.013)	−12.5
40–62	+0.271***	(0.087)	+0.598***	(0.021)	−3.8
62–125	+1.025***	(0.078)	+0.803***	(0.025)	+2.9
125–250	+0.983***	(0.067)	+1.007***	(0.037)	−0.3
250–500	+3.054***	(0.072)	+1.440***	(0.062)	+18.0
500+	+2.096***	(0.067)	+1.518***	(0.122)	+4.4

Table A11: Relation of Liquidity and Fragmentation Across Periods Decomposed by Components (continued)

	common part		idiosyncratic part		
Period (days)	β	(s.e.)	β	(s.e.)	difference z
Panel C: DEPTH, $FRAG^{lit}$					
2–3	+1.505***	(0.124)	+0.672***	(0.013)	+6.9
3–5	+1.109***	(0.131)	+0.707***	(0.014)	+3.2
5–10	+1.421***	(0.139)	+0.792***	(0.017)	+4.7
10–20	+2.605***	(0.187)	+1.064***	(0.025)	+8.6
20–40	+3.097***	(0.185)	+1.486***	(0.043)	+9.3
40–62	+5.786***	(0.221)	+2.089***	(0.070)	+17.6
62–125	+7.819***	(0.253)	+2.898***	(0.102)	+21.0
125–250	+7.462***	(0.225)	+3.649***	(0.127)	+17.6
250–500	+3.307***	(0.227)	+4.495***	(0.162)	−5.0
500+	+5.270***	(0.281)	+5.889***	(0.311)	−1.8
Panel D: DEPTH, dark					
2–3	+0.305***	(0.105)	−0.487***	(0.006)	+7.6
3–5	+0.174	(0.109)	−0.487***	(0.006)	+6.2
5–10	+0.859***	(0.101)	−0.511***	(0.006)	+13.7
10–20	+1.340***	(0.100)	−0.545***	(0.010)	+19.1
20–40	+1.709***	(0.099)	−0.648***	(0.016)	+24.3
40–62	+0.613***	(0.108)	−0.803***	(0.028)	+13.4
62–125	+0.399***	(0.095)	−0.978***	(0.040)	+14.4
125–250	+0.467***	(0.102)	−1.131***	(0.065)	+14.8
250–500	+1.288***	(0.113)	−1.045***	(0.097)	+16.7
500+	+0.947***	(0.111)	−1.220***	(0.176)	+10.7

Table A12: Engle F-Test Across Bands

The table tests the null that a coefficient is the same in all ten frequency bands of Table A10, against the alternative that it differs following Engle (1974). The test is the restricted less the unrestricted residual sum of squares, over the number of restrictions, divided by the unrestricted mean square. It is exact under his assumptions, which require homoskedastic and serially uncorrelated errors and which this panel does not satisfy. The fourth column tests the same restrictions using a Wald statistic using the covariance matrix clustered two ways, on stock and on date. The ten band regressions are fitted on orthogonal components of the same stocks and the same days, so their estimates are correlated through the clusters even though their regressors are not. *, ** and *** denote significance at the 10%, 5% and 1% levels.

Null: coefficient equal across bands	Restrictions	Engle F	Clustered χ^2
Panel A: ES			
$FRAG^{lit}$	9	11,865.4***	775.2***
dark	9	16,112.9***	2,000.8***
$FRAG^{lit}$ and *dark*	18	12,277.1***	2,586.3***
All coefficients	54	9,319.7***	8,283.4***
Panel B: DEPTH			
$FRAG^{lit}$	9	14,034.9***	1,263.6***
dark	9	3,720.3***	337.5***
$FRAG^{lit}$ and *dark*	18	7,998.0***	1,456.2***
All coefficients	54	6,659.3***	3,347.4***