Exchange-Traded Liquidity

Lars C. Larsen[1], Gyuri Venter[2] and Paul Whelan[3]

[1]Copenhagen Business School

[2]Aarhus University

[3]Chinese University of Hong Kong Business School

Abstract

This paper studies market depth in the world's most traded equity derivative market, S&P 500 E-mini futures. We show that market depth has fallen by 80% since 2012, driven by the withdrawal of liquidity providers (LPs) rather than a reduction in per-LP capacity. To shed light on this finding, we develop a model of endogenous liquidity provision that delivers testable predictions linking LP participation, volatility, and illiquidity measured by Kyle's lambda to each other and to measures of LP constraints. In line with the model's predictions, we find that tighter constraints LP numbers, and that LP participation exhibits non-monotonic relationships with volatility and illiquidity—characteristics of endogenous entry that are absent from models with exogenous market-maker populations. Overall, in this market with huge demands, we highlight that the supply of immediacy is progressively concentrated among fewer participants, contributing to higher volatility and market fragility.

Keywords: liquidity, risk-bearing capacity, liquidity supply, margin constraints

This draft: March 2026

Lars C. Larsen is at Copenhagen Business School. Gyuri Venter is at Aarhus University, School of Business and Social Sciences. Paul Whelan is at the Chinese University of Hong Kong Business School. The authors thank audiences at various conferences and seminars for their comments and suggestions. Lars C. Larsen gratefully acknowledge financial support from the Center for Financial Frictions (FRIC), grant no. DNRF102 and the Danish Finance Institute (DFI). Gyuri Venter gratefully acknowledges financial support from the Danish Finance Institute (DFI). Paul Whelan gratefully acknowledges financial support from the Hong Kong General Research Fund project number 14509725. Emails: lcl.fi@cbs.dk, venter@econ.au.dk, and paulwhelan@cuhk.edu.hk.

Liquidity—the ease with which investors can trade without moving prices—is central to asset pricing, market efficiency, and financial stability. Yet despite decades of research, we lack a clear understanding of what determines the supply of liquidity at the market level. In this paper, we exploit a previously unused feature of a publicly accessible data feed to track, at millisecond resolution, the number of distinct agents posting limit orders in the S&P 500 E-mini futures market–the world's most traded equity derivative.[1] We document that market depth has fallen by 80% since 2012, driven entirely by a decline in the number of active liquidity providers (LPs) rather than a reduction in per-LP capacity, and we develop a model of endogenous liquidity provision that rationalizes this decline and generates novel predictions for the joint dynamics of LP participation, returns, volatility, and conventional liquidity measures.

Our data span the E-mini limit order book (LOB) from 2009 to 2025, recording every update across five price levels on both the bid and ask sides. For each level, we observe the price, the number of contracts available, and, crucially, the number of distinct counterparties posting limit orders. This last variable, the LP count we denote by N, has not previously been exploited in the academic literature. Processing terabyte-scale datasets, we construct daily and intraday measures of LP participation and decompose depth into an extensive margin (the number of LPs) and an intensive margin (depth per LP). The decomposition reveals a stark fact: the 80% decline in E-mini depth between 2012 and 2025—from USD 18.4 million to USD 3.6 million at the best prices—is accounted for almost entirely by a corresponding 80% decline in N, from 52 to 11 LPs at level 1, while per-LP depth has remained roughly constant.

To understand these patterns, we develop a model in which LPs, facing participation costs and position constraints, endogenously choose whether to enter a market to provide liquidity. The model delivers a key insight: because both the number of LPs and the equilibrium price are jointly determined, the relationships among N, volatility, expected returns, and illiquidity are non-monotonic. When LP risk-bearing capacity declines—through tighter constraints or higher risk aversion—expected profits initially rise, encouraging entry, but eventually fall as aggregate risk-sharing capacity deteriorates. This generates a hump-shaped relationship between $\log N$ and the logarithm of return volatility (and between $\log N$ and log illiquidity) that distinguishes our endogenous-participation framework from models with exogenous LP numbers, under which these relationships are monotonically linear. Higher participation costs shift the equilibrium toward fewer

[1]The open interest (number of active contracts) of the E-mini is around 3 million. With an index value of 5,000 and accounting for the index multiplier of USD 50, the total value of the market is USD 750 billion.

LPs while preserving the non-monotonic shape, providing a natural explanation for the secular decline in N.

We take these predictions to the data. The model's central prediction is a hump-shaped relationship between $\log N$ and $\log \sigma$: LP participation first increases and then decreases in volatility. Estimating a quadratic specification, we find that the squared term is negative and statistically significant across all sub-samples, rejecting the linear prediction of exogenous-LP models. The turning point—the volatility level at which LP participation peaks—shifts progressively leftward over time, consistent with rising participation costs that push the equilibrium toward the downward-sloping part of the hump.

The model also pins down the relationships among illiquidity, volatility and LP participation. First, it predicts that the logarithm of illiquidity (measured by price impact or Kyle's λ) is linear in log volatility with a unit slope. In the earliest sub-period (2009–2013), we estimate a slope of 1.04 and cannot reject this prediction. The model further predicts hump-shaped relationships between $\log N$ and both $\log \lambda$. These are confirmed in the data, with the curvature becoming more pronounced in recent sub-periods as the equilibrium shifts along the curve.

A natural concern is that N and σ are jointly determined, so the regressions cannot on their own establish a causal link running from participation costs to LP entry. To address this, we exploit variation in LP constraint proxies that are plausibly exogenous to the contemporaneous entry decision: exchange margin requirements set by the CME, the He, Kelly, and Manela (2017) capital ratio of primary dealers, and G14 dealer CDS rates. Each proxy enters with the predicted sign–higher margins and CDS rates are associated with fewer LPs, while a higher capital ratio is associated with more–supporting the model's prediction that higher participation costs shift the equilibrium curves inward. Large volatility shocks–the 2018 VIX spike and the 2020 COVID-19 crash–produced persistent drops in LP participation that did not reverse even years after volatility subsided, consistent with an upward ratchet in effective participation costs.

The paper contributes to several literatures. Our analysis of endogenous participation decisions extends the canonical frameworks of Grossman and Miller (1988), Huang and Wang (2009, 2010), and Vayanos and Wang (2013), who study liquidity provision under the assumption that market makers are always present (see also Kyle, 1985; Glosten and Milgrom, 1985). We derive the equilibrium number of LPs and characterize its non-monotonic relationships with returns, volatility, and conventional liquidity measures–relationships that are absent from models with exogenous market

maker populations.

We also contribute to liquidity measurement by showing that the LP count captures information beyond traditional metrics (Kyle, 1985; Amihud, 2002; Nagel, 2012). Our findings complement the square-root impact literature, which establishes that price displacement from a metaorder of size Q scales as $\sigma\sqrt{Q/V}$, where V captures effective market depth (Bouchaud, Farmer, and Lillo, 2009; Kyle and Obizhaeva, 2016). As LPs withdraw and depth declines, this coefficient rises, implying that the same order flow generates larger price dislocations (Bouchaud, 2022).

Our work connects to the literature on high-frequency trading and market quality.[2] Finally, we relate to the literature on margin regulation and balance sheet constraints (Moore, 1966; Schwert, 1989; Hsieh and Merton, 1990; Frazzini and Pedersen, 2014; Jylhä, 2018), slow-moving capital (Duffie, 2010), and liquidity spirals (Brunnermeier and Pedersen, 2009), as well as work on the 2010 Flash Crash (Easley, López de Prado, and O'Hara, 2011, 2012; Kirilenko, Kyle, Samadi, and Tuzun, 2017).

The remainder of the paper is organized as follows. Section I presents the model of endogenous liquidity provision. Section II describes the data, market structure, and key empirical facts about depth and LP numbers. Section III tests the model predictions. Section IV concludes.

I. Theoretical framework

We formulate a simple model of endogenous liquidity provision in exchange-traded markets to capture the relationship among the number of active liquidity providers, their risk-bearing capacity, market liquidity, and volatility. We have in mind a futures market, like the S&P 500 E-mini, with futures being in zero net supply and market participants trading on margin accounts that limit their positions. Our framework builds on the canonical models of Grossman and Miller (1988), while incorporating capital constraints and participation costs as in Brunnermeier and Pedersen (2009) and Vayanos and Wang (2013).

[2]Most papers (see, e.g. Hendershott, Jones, and Menkveld, 2011, Hasbrouck and Saar, 2013, and Brogaard, Hendershott, and Riordan, 2014) find that algorithmic trading improves liquidity metrics like spreads and price discovery, though Biais and Foucault (2014) and O'Hara (2015) highlight concerns about market fragility. See also Carrion, Moyaert, Riordan, Shkilko, and Sokolov (2013), Hagströmer, Nordén, and Zhang (2014), Baron, Brogaard, Hagströmer, and Kirilenko (2015), Budish, Cramton, and Shim (2015), Foucault, Kozhan, and Tham (2015) and van Kervel and Menkveld (2018), among others.

A. Setup

There are three dates, $t = 0$, 1, and 2. The financial market consists of a riskless and a risky asset that can be traded at date 1 and pay off at date 2. The riskless asset is in elastic supply and its riskfree net return is normalized to zero. The risky asset, whose supply is normalized to zero for simplicity, pays off D units, which is normally distributed with mean $\bar{D}$ and variance σ^2. Using the riskless asset as the numeraire, we denote by p the risky asset's price at date 1.

The economy is populated by two types of agents. The first set of agents, in exogenously given mass that we normalize to 1 and have exponential utility with the coefficient of absolute risk aversion α_d, so they maximize

$$\mathrm{E}\left[\exp\left\{-\alpha_d W_{d,2}\right\}\right], \tag{1}$$

where $W_{d,2}$ is the terminal (date-2) wealth. These agents' terminal wealth consists of a starting wealth $W_{d,1}$, capital gains from trading the risky asset, and an exogenous endowment $z\left(D - \overline{D}\right)$ of the consumption good at date 2 that covaries with the payoff of the risky asset:

$$W_{d,2} = W_{d,1} + x_d\left(D - p\right) + z\left(D - \overline{D}\right), \tag{2}$$

where x_d denotes their position in the risky asset. We assume that the variable z is Gaussian with mean zero and variance σ_z^2, and is independent of D. While the endowment is received at date 2, agents learn z at date 1. Since the endowment is correlated with D, it generates a hedging demand. When, for example, $z > 0$, the endowment exposes agents to the risk that D will be low, and they hedge against this risk by selling the risky asset.[3]

The second group of agents also have exponential utility with the coefficient of absolute risk aversion α_s. They do not receive any liquidity shocks, but are subject to a constraint on the position they can take. Formally, they maximize their expected date-2 utility,

$$\mathrm{E}\left[\exp\left\{-\alpha_s W_{s,2}\right\}\right], \tag{3}$$

subject to the budget constraint

$$W_{s,2} = W_{s,1} + x_s\left(D - p\right) \tag{4}$$

[3]We assume that the endowment is perfectly correlated with D for simplicity; what matters for our analysis is that the correlation is non-zero and hence generates trade.

and the margin constraint

$$\kappa |x_s| \leq W_{s,1}, \tag{5}$$

where $W_{s,1}$ denotes the starting wealth, x_s is the position in the risky asset, and $\kappa \geq 0$ is the capital requirement associated with trading the risky asset. Constraint (5) states that in order to undertake a trade of size $|x_s|$, agents must commit $\kappa |x_s|$ of their scarce capital. Historically, margins have been set between 4% and 12% of the E-mini's contract value. Note that, after some algebra, (5) is equivalent to a Value-at-Risk constraint

$$\text{Var}\left[W_{s,2}|p\right] = \text{Var}\left[x_s D\right] = x_s^2 \sigma^2 \leq \left(\frac{W_{s,1}}{\kappa}\right)^2 \sigma^2. \tag{6}$$

To simplify equilibrium formulas, in what follows we work with this VaR constraint, (6).[4]

In equilibrium, the first group of agents initiate trades to share risk. Because they can be thought of as consuming market liquidity, we refer to them as *liquidity demanders*, and we refer to z as the *liquidity shock*. Agents in the second group accommodate the trades of liquidity demanders and supply immediacy, so we refer to them as *liquidity providers* (LPs, henceforth).

At $t = 0$, potential LPs decide whether to enter the market. Entry requires paying a fixed cost $c > 0$, interpreted as a cost of monitoring, data acquisition, or trading technology, and implies $W_{s,1} = W_s - c$. Those who do not pay the cost have $W_{s,1} = W_s$, but are not allowed to trade, $x_s \equiv 0$.[5] We denote the mass of LPs who enter by N. Thus, N is endogenous and depends on the relative value of cost c and the expected profitability of providing liquidity.

Finally, market clearing at date-1 requires that the aggregate demand of liquidity demanders and suppliers equals the zero net supply:

$$N x_s(p) + x_d(p) = 0. \tag{7}$$

[4] Constraint (6) captures VaR in terms of the date-2 riskiness of the position, conditional on information available at date 1. Alternatively, we could impose the constraint in terms of the risk of the position at date 1, given by $\text{Var}\left[x_s p\right] = x_s^2 \text{Var}\left[p\right] \leq \bar{\sigma}^2$ for some $\bar{\sigma} > 0$. This would require solving a fixed-point problem as the positional limit depends on price volatility, which in turn depends on the position (Brunnermeier and Pedersen, 2009), further complicating the problem without qualitatively impacting our results regarding the relationship of equilibrium liquidity provision, market liquidity, and volatility.

[5] See evidence about these types of costs, e.g., in Menkveld (2016), who argues that technological intermediation and high-frequency trading (HFT) have transformed market-making into a highly capital-intensive activity, where entry is limited by fixed technological and informational costs.

B. *Equilibrium*

To solve for the equilibrium of the model, we first compute the equilibrium price and quantities at date 1 given the mass of market participants; then we study LPs' entry decision at date 0.

At date 1, liquidity demanders choose holding x_1^d of the risky asset to maximize the expectation of (1). Substituting (2) into (1), after standard calculations, we obtain that the optimal demand of liquidity demanders is given by

$$x_d(p) = \frac{\overline{D} - p}{\alpha_d \sigma^2} - z. \tag{8}$$

This expression has two components: a downward-sloping demand curve, according to which they buy the asset as long as the price is below the fundamental value and (short) sell otherwise, and a term influenced by the liquidity shock z. When, for example, z is positive and large, liquidity demanders are willing to sell because their endowment is positively correlated with the asset.

Conditional on having entered, each LP chooses a position x_s in the risky asset to maximize the certainty equivalent

$$CE[W_{s,2}] \equiv W_{s,1} + x_s\left(\overline{D} - p\right) - \frac{1}{2}\alpha_s \sigma^2 (x_s)^2, \tag{9}$$

subject to the margin/VaR constraint (6). After standard calculations, the optimal position is

$$x_s(p) = \frac{\overline{D} - p}{\left(\alpha_s + \frac{1}{2}\eta_s\right)\sigma^2}, \tag{10}$$

where η_s denotes the Lagrange multiplier of the constraint (6).

Substituting (8) and (10) into the market-clearing condition (7), we obtain the equilibrium price

$$p = \overline{D} - \frac{\sigma^2}{N\frac{1}{\alpha_s + \frac{1}{2}\eta_s} + \frac{1}{\alpha_d}} z. \tag{11}$$

We find that since LPs are risk averse and/or are subject to margin/VaR constraints, the equilibrium price p is decreasing in the liquidity shock z. For a positive z, for example, liquidity demanders want to sell, and the price must drop so that LPs are willing to buy the asset. The risk premium LPs demand for buying the asset is higher when they are more risk averse (higher α_s), their constraint is tighter (higher η_s), and the asset is riskier (higher σ^2). Equilibrium positions of liquidity demanders and suppliers, x_d and x_s, then follow from combining (8) and (10) with (11).

From the above expressions, we note that LPs' risk aversion and margin/VaR constraints impact

equilibrium prices the same way as they both limit LPs' ability to provide liquidity. As a result, to simplify our calculations, in what follows we focus on the impact of risk aversion α_s on prices, returns, and quantities, and set $\eta_s = 0$ in the formulas.[6]

At $t = 0$, potential LPs compare the expected utility from entering the market to that of staying out. Combining (9)-(11), an LP enters as long as

$$\mathbb{E}[-\exp\{-\alpha_s[W_s - c + \frac{1}{2}\frac{\alpha_s\sigma^2}{(N+\frac{\alpha_s}{\alpha_d})^2}z^2]\}] \geq -\exp\{-\alpha_s W_s\}, \tag{12}$$

which can be rearranged to obtain an explicit condition for participation. Intuitively, entry is more likely when expected trading profits–which increase with the magnitude of $|z|$–are high or when participation costs c are low. LPs' risk aversion, however, has an ambiguous impact on their expected profits: higher α_s imply that LPs are less willing to participate in risk sharing, meaning lower equilibrium positions and higher required returns as shown by (11). Combining the entry condition with the distribution of shocks yields a closed-form expression for the equilibrium mass of active LPs as follows:

Theorem 1. (Equilibrium LP number and risk aversion) *For $c = 0$, liquidity provision is unlimited, $N \to \infty$. When $0 < \alpha_s c < \bar{C} \equiv \frac{1}{2}\log(1 + \alpha_d^2\sigma^2\sigma_z^2)$, the equilibrium number of LPs is finite and given by*

$$N = \sqrt{\frac{\alpha_s^2\sigma^2\sigma_z^2}{\exp\{2\alpha_s c\} - 1}} - \frac{\alpha_s}{\alpha_d} > 0. \tag{13}$$

In addition, there exists $\bar{\alpha}_s > 0$ such that $dN/d\alpha_s > 0$ when $0 < \alpha_s < \bar{\alpha}_s$ but $dN/d\alpha_s < 0$ when $\alpha_s > \bar{\alpha}_s$. Finally, if $c \geq \bar{c}$, we have $N = 0$.

Theorem 1 shows that as long as the participation cost c is positive but not too high, the equilibrium number of LPs is hump-shaped in α_s: it is increasing for small values of α_s and decreasing otherwise. In particular, for small α_s the positive impact of higher risk aversion on the per-unit expected profit dominates its negative impact on the quantity of the risky asset LPs are willing to take on their books. At the same time, higher risk aversion eventually deters entry as each LP would only offer to trade a small quantity and earn low profits.

[6] In the Appendix, we also solve a model in which LPs are risk neutral but are subject to the margin constraint. We confirm that higher risk aversion and tighter constraints have qualitatively similar impacts on the variables of interest, namely on the equilibrium number of LPs, returns, volatility, and liquidity measures. Note that when $c > 0$ and we abstract away from the constraint, LPs must be risk averse: Risk neutrality leads to the price equalling the expected payoff, leading to zero expected trading profits, in which case no LP is paying the entry cost and there is no equilibrium of the full model.

Next, we provide testable predictions of the model that relate the number of LPs with first and second moments of asset returns and liquidity measures in the market while assuming that all of these are driven by variation in LPs' risk aversion (or constraint tightness). We derive formal relationships under the assumption $\alpha_d \to \infty$, which means liquidity demanders submit price-inelastic demands that hedge the liquidity shock perfectly, and signed volume from them equals $-z$. This assumption simplifies equilibrium objects without qualitatively impacting the results.[7] Note also that, by definition, N and measures of volatility and liquidity are subject to a zero lower bound, so their relationships can flatten out for extreme observations, biasing our estimates in simple linear or quadratic regressions. To overcome this problem, in our analysis we will work with the logarithm of equilibrium objects of interest. Under the assumption of $\alpha_d \to \infty$, (13) yields

$$\log N = \log \sigma\sigma_z + \log \alpha_s - \frac{1}{2} \log \left(\exp \left\{ 2\alpha_s c \right\} - 1 \right). \tag{14}$$

Let us set a hypothetical date-0 price to be the expected payoff of the asset, $\overline{D}$, which allows us to define *realized returns* calculated as those between dates 0 and 1, $R_1 = p - \overline{D}$, and *expected returns* as those between dates 1 and 2, conditional on the price or the liquidity shock, $\mathrm{E}\left[R_2|p\right] = \mathrm{E}[R_2|z] = \overline{D} - p = -R_1$.[8] Combining (11) and (13), equilibrium returns are

$$R_1 = -\mathrm{E}\left[R_2|p\right] = -\frac{\sigma\sqrt{\exp\left\{2\alpha_s c\right\} - 1}}{\sigma_z} z, \tag{15}$$

and return or price volatility are given by

$$\log \sigma_R \equiv \log \sqrt{\mathrm{Var}\left[p\right]} = \log \sqrt{\mathrm{Var}\left[R_1\right]} = \log \sigma + \log \left(\exp \left\{ 2\alpha_s c \right\} - 1 \right). \tag{16}$$

Following Vayanos and Wang (2013), we also explore the impact of the number of LPs on two measures of illiquidity, both based on the price impact of the liquidity demanders' trades at date 1. The first measure is *price impact*, and can be either measured as Kyle (1985)'s lambda–i.e., the impact of the liquidity shock z on the price in (11): $\lambda_1 \equiv -\frac{dp}{dz} = -\frac{dR_1}{dz}$, or as the coefficient of a regression of the price change between dates 0 and 1 on the signed volume of liquidity demanders–i.e.,

[7] Note that the absence of the α_s/α_d term in (13) lowers the value and slope of N as a function of α_s but does not change the convexity of the relationship. In addition, $\bar{c} \to \infty$, so $N > 0$ for all $c, \alpha_s > 0$.

[8] Vayanos and Wang (2013) solve for the date-0 equilibrium price at which agents would trade among each other before learning which of them are liquidity demanders or suppliers. As they show, liquidity risk impacts the date-0 price by impacting the multiplier of the net supply of the asset, but since we work with futures that are in zero net supply, our equilibrium date-0 price would be the expected payoff of the asset, $\overline{D}$.

$\lambda_2 \equiv -\operatorname{Cov}[R_1, x_d]/\operatorname{Var}[x_d]$. From (8) and (11), in equilibrium these two price impact measures are

$$\log \lambda_1 = \log \lambda_2 = \log \frac{\sigma}{\sigma_z} + \log\left(\exp\{2\alpha_s c\} - 1\right) \equiv \log \lambda. \tag{17}$$

The second measure of illiquidity is based on *return reversal*, measured as the autocovariance of returns. Liquidity demanders' trades at date 1 cause the price to deviate from fundamental value, while the two coincide at date 2. Therefore, price changes exhibit negative autocovariance, and more so when trades have large price impact. We use minus unconditional autocovariance to measure reversal, $\gamma_1 \equiv -\operatorname{Cov}[R_2, R_1]$, and in equilibrium, we find that

$$\log \gamma_1 = 2\log \sigma + 2\log\left(\exp\{2\alpha_s c\} - 1\right) = 2\log \sigma_R, \tag{18}$$

i.e., price reversal is higher when LPs are more risk-averse and the asset is riskier.[9]

From here we obtain the following results:

Proposition 1. (LP number and volatility) *Assuming the variation in the number of LPs and volatility is driven by variation in risk aversion, there exists $\bar{\sigma} > 0$ such that $d\log N/d\log\sigma_R > 0$ when $0 < \sigma_R < \bar{\sigma}$ but $d\log N/d\log\sigma_R < 0$ when $\sigma_R > \bar{\sigma}$.*

As their risk aversion increases, keeping everything else equal, LPs are less willing to offset liquidity shocks, leading to higher price volatility, and require higher premia for buying the asset from liquidity demanders. This in turn increases expected profits and encourages entry from additional LPs, which lowers the expected profits of incumbents. In equilibrium, we find that while equilibrium volatility increases in α_s, the number of LPs first increases and then decreases in α_s, and overall $\log N$ exhibits a hump-shaped relationship in relation to log volatility.

Combining (16)-(18), we see that higher risk aversion leads to lower liquidity measured by higher price impact, $d\log\lambda/d\alpha_s > 0$, and higher return reversal, $d\log\gamma_1/d\alpha_s > 0$. Therefore, we can relate volatility and liquidity:

Proposition 2. (Volatility and liquidity) *There are positive linear relationships between price impact and volatility, $d\log\lambda/d\log\sigma_R > 0$, and between reversal and volatility, $d\log\gamma_1/d\log\sigma_R > 0$.*

In addition, Propositions 1 and 2 together imply:

[9] Looking at the reversal measure conditional on the liquidity shock, $\gamma_2 \equiv -\operatorname{Cov}[R_2, R_1|z]$, or at return autocorrelation, would have the same qualitative properties as γ_1.

Proposition 3. (LP number and liquidity) *There exists* $\bar{\lambda} > 0$ *such that* $d \log N / d \log \lambda > 0$ *when* $0 < \lambda < \bar{\lambda}$ *but* $d \log N / d \log \lambda < 0$ *when* $\lambda > \bar{\lambda}$.

Expected returns compensate risk-averse LPs for the risk they bear in equilibrium, and the required risk premium increases proportionally to the number of units that they need to hold to provide liquidity, with the coefficient being equal to λ as seen from (11). Therefore, we find that there is a hump-shaped relationship between the equilibrium LP number and the impact of signed volume on contemporaneous returns, $\beta_{R_1,z} \equiv Cov[R_1, z]/\sigma_z^2$. In addition, we can think about the equilibrium relationship between the LP number and absolute returns unconditionally, and this relationship is also hump-shaped. Formally:

Proposition 4. (LP number and returns) *There exists* $\bar{\beta} > 0$ *such that* $d \log N / d\beta_{R_1,z} > 0$ *when* $0 < \beta_{R_1,z} < \bar{\beta}$ *but* $d \log N / d\beta_{R_1,z} < 0$ *when* $\beta_{R_1,z} > \bar{\beta}$.
In addition, holding $z > 0$ *constant, there exists* $|\bar{R}| > 0$ *such that* $d \log N / d|R| > 0$ *when* $0 < |R| < \bar{R}$ *but* $d \log N / d|R| < 0$ *when* $|R| > \bar{R}$.

When the liquidity shock is positive, $z > 0$, liquidity demanders want to sell the asset, which pushes the equilibrium price below the fundamental value. Since LPs are risk averse, they require a risk premium to hold the asset, and (15) shows that this risk premium needs to be higher (R_1 is more negative) for more risk-averse LPs. Since from (13) we know that the number of LPs is hump-shaped in α_s, we conclude that for low levels of risk aversion, when $z > 0$, both $\mathrm{E}[R_2|p]$ and N are increasing in α_s, whereas for high level of risk aversion, $\mathrm{E}[R_2|p]$ is increasing but N is decreasing in α_s.

Next, we examine how the relationship between return moments or liquidity and the number of LPs changes if the participation cost c increases. We think about higher c as higher cost of monitoring, data acquisition, or trading technology; in particular, we posit that these costs have significantly increased over time as the competition between various LPs such as trading desks, hedge funds, and high-frequency traders has increased. We obtain the following results:

Proposition 5. (LP number-volatility and LP number-liquidity curve shifts) *Assume that* $c > 0$. *An increase in* c *while keeping* $c \cdot \alpha_s$ *constant lowers* N *while keeping volatility* (σ_R) *and liquidity* (λ *and* γ_1) *constant. As a result, when participation gets more costly, (i) the* $\log N - \log \sigma_R$ *curve shifts "downwards," and the same level of* σ_R *corresponds to a lower equilibrium LP number*

N; (ii) the $\log N - \log \lambda$ and $\log N - \log \gamma_i$ curves shift "downwards," and the same level of liquidity correspond to a lower equilibrium LP number N.

Proposition 5 shows that rising data and infrastructure costs in modern electronic markets can reduce equilibrium liquidity even as technology advances. As we show in Section III, these comparative statics align with empirical evidence from exchange-traded derivatives such as the E-mini S&P 500, where market depth and the number of active liquidity providers have declined substantially despite increased quote activity. The framework thus links structural shifts in participation and risk-bearing capacity to observed patterns in returns, volatility, liquidity, and market fragility.

C. Discussion

Before we turn our attention to the empirical analysis, we discuss our assumptions and the robustness of the results, and relate our findings to those in the previous literature.

The solution of our model and hence its testable implications hinge on a couple of assumptions. In the model presented above, participation costs paid by potential liquidity suppliers ensures that we obtain an endogenously determined number of LPs as opposed to this number being exogenously given. If we set $c = 0$, the maximum possible number of risk-neutral LPs would enter the market because, as (12) shows, they would have positive expected profits from providing liquidity. Having $c > 0$ participation cost thus provides us with a finite N and equilibrium relationships between N and expected and realized returns, N and volatility, and N and liquidity, and the mechanism behind this relationship is via the impact of risk-bearing capacity α_s on these variables.

The assumption that LPs are risk averse but not subject to margin or VaR constraints has no qualitative impact on our results. In the Online Appendix we present a version of the model that allows both for risk aversion and these constraints. Its main implication is that LPs' leverage constraints only occasionally bind, depending on the size of the liquidity shock, which further complicates the expressions for equilibrium prices, quantities and moments, without providing additional insights about their dynamics.

The model provides a tractable way to interpret recent changes in exchange-traded liquidity. Following Menkveld (2016), technological intermediation and high-frequency trading (HFT) have transformed market-making into a highly capital-intensive activity, where entry is limited by fixed technological and informational costs. In equilibrium, only a subset of potential LPs remain active. The mechanism echoes Duffie (2023), who emphasizes that systemic liquidity fragility can arise

even in competitive markets when risk-bearing capacity and/or balance-sheet constraints tighten collectively.

II. Data and Market Structure

Like forwards, a futures contract specifies the exchange of an asset for a specified price at a specified future date. Unlike forwards, futures are standardized contracts that trade on exchanges, making them more liquid due to larger trading volume of each specific contract. Futures are also margined, making them significantly less exposed to credit risk.

The first S&P 500 futures contract was introduced in 1982 and quickly became popular as it allowed traders to easily speculate on or hedge the equity market. The E-mini futures started trading in 1997 and became the most traded equity contract by 2004.

Futures contracts trade on central exchanges, where market participants post either limit orders or market orders. A limit order is an order to buy or sell at a specific price–the bid and ask price, respectively. Limit orders thus supply immediacy and liquidity to the market and are registered in the limit order book (LOB). In contrast, a market order is instantly executed at the best available price in the LOB and therefore consumes liquidity.

If a large number of agents are providing liquidity by posting limit orders, the bid–ask spread will be low due to market competition, making the spread a natural measure of liquidity. In the extreme, in large and competitive futures markets, competition drives the bid-ask spread to the minimum tick size–the minimum possible price change in the market–which renders the bid-ask spread an uninformative measure of liquidity.

Instead, we focus on market depth or *dollar volume*, defined as the number of contracts available in the LOB $\times$ the price of the contract. Naturally, a high market depth implies that prices stay relatively flat even when trading volume is high. The dollar volume for a given level of the LOB is computed as

$$DVol_j = C_j \cdot P_j \cdot m, \tag{19}$$

where $j = 1, \ldots, 5$ is the order book level, C_j is the number of contracts available (i.e., the number of posted limit orders), P_j is the price level, and m is the index multiplier. Since we know the total

number of limit order providers, we rewrite (19) as

$$DVol_j = N_j \cdot \frac{C_j}{N_j} \cdot P_j \cdot m, \tag{20}$$

where N_j is the number of LPs that are posting a limit order and C_j/N_j is the average number of contracts offered per LP. This decomposition allows us to understand whether it is the total number of LPs that drive (il)liquidity.

A. Data

Our primary data source is the market depth of the S&P 500 E-mini futures contract–the world's most traded derivative–traded on CME Globex, an electronic exchange where market participants can submit limit orders and market orders. The data are obtained from Refinitiv Data (previously Thomson Reuters) and, for levels 1 to 5 of the limit order book (LOB), contain (i) the number of contracts (limit orders) available to trade, C, (ii) the number of LPs (market participants that are currently posting limit orders), N, and (iii) the quoted price, P. The E-mini depth data is available from January, 2009 and we end our sample period in December 2024.[10]

The raw data come as tick data, meaning that every single update to the limit order book (LOB) is registered as a new observation. Updates to the LOB can be a market order that will reduce the depth of the LOB, it can be a LP posting new limit order (i.e., quotes) which increases the depth, or it can be a LP cancelling existing quotes. We choose to sample the tick data, which contain millions of observations per day, both at a daily frequency and at a 5-minute frequency, giving us 288 observation per trading day. This enables us to both study the time series and the intraday behaviour of liquidity in high frequency. For each 5-minute interval, C, N and P are computed as averages of all observations in a given interval.

We note that the data do not show the number of contracts offered by individual LPs, so we cannot identify individual LPs or how LPs change their positions. For example, if there are 10 available contracts to buy from three LPs at an ask price of 100 at level 1 of the LOB, then we know that the three LPs are posting at least one contract each, but we do not know if they offer 1,1, and 8 contracts, respectively, or 3, 3 and 4 contracts.

We classify changes in the limit order book into limit orders, market orders, and cancellations

[10]The best bid and ask prices are referred to as level 1 of the LOB, the second best bid and ask prices are referred to as level 2 of the LOB, etc.

using a tick-by-tick algorithm that processes every update across all five levels on both sides of the book. The algorithm and its detailed description are presented in Section A.3 of the Online Appendix.

B. Liquidity in the S&P 500 E-mini Futures Market

We start by documenting how liquidity of the E-mini contract has evolved over time. Figure 1 plots the time series of liquidity measured by market depth (dollar volume). The daily dollar volume is computed as the average of all 5-minute observations during the trading day. Overall, depth is highly time-varying and exhibits downwards jumps that are followed by prolonged period where the depth slowly rebounds.[11] Notable examples of jumps are the financial crisis (the September, 16, 2008 event is not included in our sample period, but the rebound in depth is evident), the 2010 flash crash, the 2011 European sovereign debt crisis, the 2015–2016 stock market selloff, the February 5 2018 VIX spike (volmageddon) and the 2020 COVID-19 stock market crash. Surprisingly, the two latter shocks appear to have had persistent effects as depth has not rebounded to its previous levels, even after more than six years in the case of the 2018 event.

As described in (20), market depth can be decomposed into the number of LPs and the depth offered per LP. This allows us to test what drives changes in market depth: does the number of LPs, N, change, or do each individual LP change the number of contracts that they are willing to trade. Figure 2 Panel (a) displays the time series of active LPs and shows that there is indeed fewer LPs posting limit orders over time. Figure 2 Panel (b) plots the time series of depth offered per LP. While the dollar volume offered per LP is time-varying and drops during crises, we do not observe a downward trend as we do for the LP number. In a simple linear regression of the daily time series, the LP number explains 79% of variation in market depth, while depth per LP explains 41%.[12] Therefore, the long-term lower depth level observed in figure 1 is primarily driven by a lower number of LPs. This empirically supports our emphasis on using the participation decision of LPs as a liquidity measure.

Considering the overall trend in market depth, the average dollar volume available to trade at both the best bid price and best ask price was USD 18.4 million in 2012 and USD 3.6 million in

[11] This holds for all individual levels of the LOB and for both the number of contract available and the dollar volume available.

[12] In a multiple linear regression, the LP number and depth per LP explains 96% of variation in total depth. We note that because total depth equals N times depth-per-LP by construction, the two regressors are not independent: any change in N mechanically affects the ratio. The R^2 comparison is therefore illustrative of the relative importance of the extensive margin (number of LPs) versus the intensive margin (per-LP provision), rather than a test of orthogonal drivers.

2022. This corresponds to a drop of 80% and similar drops are seen at all price levels of the LOB and at all times during the trading day. The decline in depth is driven by the drop in LP numbers, where there in 2012 was an average of 52 LPs per day at level 1 of the LOB, and 11 in 2022. Likewise, the average LP number across level 1 to 5 was 100 in 2012 and 21 in 2022–again a drop of 80%.[13]

[Insert Figures 1 and 2 here]

A comparison of the active LP count N with the number of CFTC-registered dealers–and a detailed analysis of intraday depth patterns–can be found in Section A.4 of the Online Appendix.

III. Testing Model Predictions

We now test the predictions of our model. We measure liquidity as the number of LPs present in the market in accordance with our model framework; however, all results hold equally well when using market depth (dollar volume), because the variation in the latter is predominantly caused by the variation in the number of LPs. We discuss traditional liquidity measures (bid–ask spread, Kyle's λ, and Amihud's illiquidity measure) in the Online Appendix (Section A.5).

A. Prediction 1: LP Number and Volatility

Proposition 1 predicts a hump-shaped relationship between $\log N$ and $\log \sigma_R$: the LP number first increases and then decreases in volatility. We measure volatility using daily realized volatility, RV, computed from logarithmic returns sampled at a 5-minute frequency throughout the trading day.[14] We test this prediction by estimating

$$\log N_{t,h} = \alpha + \beta_1 \log \sigma_{t,h} + \beta_2 (\log \sigma_{t,h})^2 + \epsilon_t, \tag{21}$$

where $N_{t,h}$ is the number of LPs during trading period h on day t, and $\sigma_{t,h}$ is the annualized realized volatility computed from 1-minute log returns.

Figure 3 plots the average daily number of LPs, N, against daily realized volatility, split by sub-period. The figure reveals two key features: a strong negative relationship between the LP number

[13] Our results complement Nagel (2012), who documents a secular decline in equity market liquidity measured by return reversals, whereas we focus on market depth and the number of LPs in the E-mini market.

[14] All following results on volatility hold equally well when using the VIX as the measure of volatility.

and volatility, and a pronounced downward shift in the N–σ curve over time. The N-to-volatility ratio peaked during 2009–2013 and has been at its lowest from 2018 onwards. In Panel (b), the quadratic regressions in logs have R^2 values around 0.7 and t-statistics exceeding 10.

[Insert Figure 3 here]

Table I reports estimates of (21) for three sub-periods. Across all specifications, a strong and negative relationship emerges between the LP number and volatility. The quadratic specification implies a hump-shaped relationship with a turning point at $\log \sigma^* = -\hat{\beta}_1/(2\hat{\beta}_2)$, above which $\log N$ decreases in $\log \sigma$. Across sub-periods, the turning point shifts progressively leftward: $\log \sigma^* = 0.68$ in 2009–2013, $\log \sigma^* = 1.26$ in 2013–2018, and $\log \sigma^* = -1.06$ in 2018–2024. In the most recent sub-period, the turning point falls below the range of observed volatilities, meaning the relationship is effectively monotonically decreasing–consistent with LPs withdrawing more aggressively at all volatility levels, in line with higher participation costs shifting the equilibrium.

The quadratic specification also provides a direct test of whether LP participation is endogenous. Under a simpler model in which N is exogenous–or determined independently of volatility–$\log N$ would be at most linear in $\log \sigma$, implying $\beta_2 = 0$. Our model, in contrast, predicts a non-zero quadratic coefficient because the entry decision creates a non-linear feedback between volatility and participation. In the CTC specification, $\hat{\beta}_2$ is negative and statistically significant in every sub-period: -0.14 ($t = -2.33$) for 2009–2013, -0.23 ($t = -3.83$) for 2013–2018, and -0.08 ($t = -2.00$) for 2018–2024. The rejection of $\beta_2 = 0$ is consistent across all hourly trading periods and aggregation windows in Table I, supporting the endogenous-entry mechanism in our model.

[Insert Table I here]

B. Prediction 2: Illiquidity and Volatility

Proposition 2 predicts a positive linear relationship between illiquidity and volatility: $\log \lambda = \log \sigma - \log \sigma_z$, implying a unit slope and an intercept equal to $-\log \sigma_z$. We test this by estimating

$$\log \lambda_t = a + b \cdot \log \sigma_t + \epsilon_t, \tag{22}$$

where λ_t is the daily average of hourly Kyle's λ during U.S. trading hours and σ_t is the annualized daily realized volatility from 5-minute log returns.

Figure 4 plots the daily cross-section of $\log \lambda$ and $\log \sigma$ split by sub-period, and Table II reports the OLS regressions. In the earliest sub-period (2009–2013), the estimated slope is $\hat{b} = 1.04$ with $t(b = 1) = 0.56$: we cannot reject the model's prediction of a unit slope. In the later sub-periods, the slope attenuates to 0.60 and 0.64, significantly below one, potentially reflecting structural changes in market microstructure or measurement error. The intercept shifts upward from -20.6 to -19.4 across sub-periods, consistent with a decline in order flow volatility σ_z over time (since $-\log \sigma_z$ increases as σ_z falls).

[Insert Figure 4 and Table II here]

C. *Prediction 3: LP Number and Illiquidity*

Proposition 3 predicts a hump-shaped relationship between $\log N$ and $\log \lambda$. We test this by estimating

$$\log N_t = a + b_1 \cdot \log \lambda_t + b_2 \cdot (\log \lambda_t)^2 + \epsilon_t. \tag{23}$$

Figure 5 plots this relationship by sub-period, and Table III reports the quadratic regressions. The quadratic coefficient is negative and statistically significant in the most recent sub-period: $\hat{b}_2 = -0.26$ ($t = -4.98$) for 2018–2024, confirming the hump-shaped relationship. In earlier sub-periods, the quadratic term is weaker ($t = 1.46$ for 2009–2013 and $t = -1.85$ for 2013–2018), suggesting that the curvature has become more pronounced as participation costs increased. The turning points shift across sub-periods, consistent with the equilibrium moving along the curve as market conditions change.

[Insert Figure 5 and Table III here]

D. *Prediction 4: LP Number and Price Impact*

The model predicts a hump-shaped relationship between $\log N$ and price impact β. We estimate daily price impact by running hourly regressions of 1-minute log returns on 1-minute order flow during U.S. trading hours, then averaging across hours. We test this prediction by estimating

$$\log N_t = a + b_1 \cdot \log \hat{\beta}_t + b_2 \cdot (\log \hat{\beta}_t)^2 + \epsilon_t. \tag{24}$$

Figure 6 plots the resulting scatter of $\log N$ versus $\log \hat{\beta}$ by sub-period, and Table IV reports the quadratic regressions. In the first two sub-periods, the quadratic coefficient is negative and highly significant: $\hat{b}_2 = -0.19$ ($t = -4.86$) for 2009–2013 and $\hat{b}_2 = -0.23$ ($t = -5.03$) for 2013–2018, confirming the predicted hump shape. In the most recent sub-period (2018–2024), the relationship is approximately linear ($\hat{b}_2 = -0.02$, $t = -0.99$), with a strong negative slope ($R^2 = 79\%$), indicating that higher price impact is uniformly associated with fewer LPs. This is consistent with the equilibrium having shifted to the downward-sloping portion of the hump, where participation costs are high enough that the relationship is effectively monotone.

[Insert Figure 6 and Table IV here]

E. Prediction 5: Participation Costs and Curve Shifts

Proposition 5 predicts that higher participation costs shift the $\log N$–$\log \sigma$ and $\log N$–$\log \lambda$ curves inwards: the same level of volatility or illiquidity corresponds to a lower equilibrium LP number. A natural concern with the preceding results is that the LP number N and volatility σ are jointly determined in equilibrium, so the regressions in Table I cannot, on their own, establish a causal link running from participation costs to LP entry. We address this by exploiting variation in LP constraint proxies that are plausibly exogenous to the contemporaneous LP participation decision. These proxies are not instruments in a formal 2SLS sense; rather, they provide reduced-form evidence that LP participation responds to observable constraint variables whose variation is largely determined outside the day-to-day LP entry decision.

Assuming that higher price volatility leads to tighter margins, our model suggests that the resulting feedback effect further amplifies the negative relationship between volatility and liquidity supply. If new capital then enters the market slowly (Duffie, 2010), we would expect that after a period of high volatility, the equilibrium LP number drops and recovers slowly. Higher participation costs can therefore explain why the LP number have dropped in recent years even though volatility is now low.[15]

We consider three constraint proxies, each capturing a distinct and plausibly exogenous channel

[15] Regulatory developments may have contributed to higher participation costs during our sample. In November 2015, the CFTC proposed Regulation Automated Trading (Regulation AT; 80 FR 78824), which would have required previously unregistered proprietary algorithmic trading firms–including many automated liquidity providers in futures markets–to register as floor traders, implement prescribed pre-trade risk controls, and retain source code for regulatory inspection. Although the rule was never finalized and was formally withdrawn in 2020 (85 FR 42755), the multi-year period of regulatory uncertainty may have raised perceived entry costs for automated market makers.

through which LP capacity may be affected, and test the prediction by estimating

$$\Delta N_t = \alpha + \beta \Delta D_t^i + \epsilon_t \qquad \text{and} \qquad \Delta \sigma_t = \alpha + \beta \Delta D_t^i + \epsilon_t, \tag{25}$$

where D_t^i denotes constraint proxy i and all variables are in differences and z-scored.

We cannot measure LP wealth directly, but margins are observable at all times. The futures margin requirement is the amount of cash that a market participant must deposit and keep on hand with her broker when holding a futures position. Futures margins are typically at 3–12% of the notional value of the contract, and they generally go up when price volatility increases, but the exact margin size is proprietary to the CME.[16] Crucially, margin levels are set by the exchange clearing house based on its own risk models, not by LPs themselves, making them a plausibly exogenous shifter of LP capacity. Figure A.8 depicts the time series of margins both in U.S. dollars and in percent of the E-mini contract value. We observe a jump in margins close to 100% during the COVID-19 stock market crash where volatility also was at extreme levels. This is in line with standard clearing house policy of adjusting margins relative to risk as measured through volatility, and testing for this, we find that volatility explains next day's margins. What is surprising, is that margins have remained at elevated levels since 2020 and this could potentially explain the persistent decrease in market depth, but this is in line with Abruzzo and Park (2016)) who also find that the CME is quick to increase margins in response to volatility shocks but slow to decrease them.

Table V confirms that the margin percent explains liquidity measured as the number of LPs–also when controlling for volatility, and, indeed, we see that higher margins are associated with fewer LPs. Specifically, a margin increase of 1 percentage point is associated with a drop of 14 in the expected LP number, all else equal. Additionally, Panel (b) of Table V confirms the predicted positive relationship between margins and volatility. Since margins have remained elevated since 2020, they remain as a natural candidate for explaining the E-mini depth dry-up. However, we note that according to the model, other potential increases to LPs' costs can also explain the low number of LPs and depth.

[Insert Figure A.8 here]

Second, to proxy for LPs' wealth we use the capital ratio of primary dealers as proposed by He, Kelly, and Manela (2017). Recall the model prediction that lower LP wealth leads to lower liquidity

[16] See www.cmegroup.com/education/courses/introduction-to-futures/margin-know-what-is-needed.html

and higher volatility. Thus, a higher capital ratio should increase the LP number and decrease volatility. We cannot identify the E-mini futures LPs, but the CFTC states that registered LPs consist of large U.S. and non-U.S. banks. These banks are active as LPs in most large markets, and we therefore use the He, Kelly, and Manela (2017) capital ratio as a proxy for the capital ratio of E-mini LPs. The He, Kelly, and Manela (2017) capital ratio is available at a daily frequency and is based on the list of NY Fed primary dealers. Because it reflects aggregate financial sector health–driven by bank asset values and leverage ratios that are slow-moving relative to intraday LP entry decisions–it provides a second plausibly exogenous source of variation in LP capacity. Regression (ii) in Panel (a) of Table V confirms that the capital ratio and the LP number are positively correlated, and Panel (b) shows that a higher capital ratio implies lower volatility as predicted by the model.

Third, we proxy for the LPs' risk-bearing constraint with the average CDS rates of G14 dealers, as the swap rate should increase when dealers are financially constrained. The CDS rate captures dealer credit risk and reflects funding conditions that are determined by bond market participants and credit analysts, not by futures market LPs directly.[17] Again, we find that higher constraints (higher swap rate) imply fewer LPs and higher volatility.

Column (iv) of Table V includes all three constraint proxies in a multiple regression. The margin percentage and the capital ratio remain significant, while the CDS rate does not, likely because dealer credit risk is partially captured by the other two proxies. Taken together, the results in Table V provide evidence that LP participation responds to plausibly exogenous variation in constraints, supporting the model's prediction that higher participation costs reduce the equilibrium number of LPs and increase volatility.

[Insert Table V here]

One potential explanation for the decline in depth and LP participation is that the demand for liquidity is simply lower. We examine this hypothesis using open interest data in Section A.4 of the Online Appendix and find that it is not supported: while open interest has declined modestly since 2018, the drop is far smaller than the 80% decline in depth, and the open-interest-to-depth ratio has fallen five-fold.

[17] The time series for the LP constraint proxies can be seen in Figure A.9 in the Online Appendix.

IV. Conclusion

Market depth in the S&P 500 E-mini futures–the world's most traded derivative–has fallen by 80% since 2012, driven entirely by a decline in the number of active liquidity providers rather than a reduction in per-LP capacity. We develop a model of endogenous liquidity provision that rationalizes this decline and generates five testable predictions linking LP numbers, volatility, illiquidity, price impact, and participation costs. The data confirm all five: the LP number exhibits the predicted hump-shaped relationships with volatility, illiquidity, and price impact in logs, while illiquidity and volatility are linearly related with a near-unit slope in the early sample. Exogenous variation in LP constraint proxies shifts these equilibrium relationships inward, and the progressive leftward shift of turning points across sub-periods is consistent with rising participation costs as the primary driver of the secular decline.

Large volatility events–the 2018 VIX spike and the 2020 COVID-19 crash–produced persistent drops in LP participation that did not reverse even years after volatility subsided, consistent with an upward ratchet in effective participation costs. Notably, the decline in depth has occurred despite only modest falls in open interest, suggesting that the supply of immediacy has contracted independently of its demand.

Whether this transformation represents a welfare loss depends on the source of higher participation costs. If costs reflect genuine frictions–regulatory burdens, elevated margins, or information acquisition–the decline in LP numbers reduces risk sharing. If instead market-making capacity has consolidated into fewer but more efficient firms, the welfare implications are ambiguous and hinge on the trade-off between normal-period efficiency and stress-period fragility. Resolving this question is an important direction for future work. In either case, the E-mini market's lack of designated market makers–in contrast to, for example, U.S. Treasury futures–leaves it without a structural safeguard against the complete withdrawal of voluntary liquidity providers.

References

Abruzzo, Nicholas, and Yoontae Park, 2016, Margin setting, margin changes, and the cost of hedging, *Journal of Futures Markets* 36, 1121–1157.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Baron, Matthew, Jonathan Brogaard, Björn Hagströmer, and Andrei Kirilenko, 2015, Risk and return in high-frequency trading, *Journal of Financial Economics* 116, 271–288.

Biais, Bruno, and Thierry Foucault, 2014, Hft and market quality, *Bankers, Markets & Investors* 128, 5–19.

Bouchaud, Jean-Philippe, 2022, The inelastic market hypothesis: A microstructural interpretation, *Quantitative Finance* 22, 1785–1795.

——— , J. Doyne Farmer, and Fabrizio Lillo, 2009, How markets slowly digest changes in supply and demand, in Thorsten Hens, and Klaus Reiner Schenk-Hoppé, ed.: *Handbook of Financial Markets: Dynamics and Evolution* . pp. 57–160 (Elsevier).

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-frequency trading and price discovery, *The Review of Financial Studies* 27, 2267–2306.

Brunnermeier, Markus K, and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *The Review of Financial Studies* 22, 2201–2238.

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

Carrion, Allen, Thierry Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov, 2013, High-frequency trading and the new market makers, *Journal of Financial Markets* 16, 712–740.

Deuskar, Prachi, and Timothy C Johnson, 2011, Market liquidity and flow-driven risk, *The Review of Financial Studies* 24, 721–753.

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *The Journal of Finance* 65, 1237–1267.

——— , 2023, Financial regulatory reform after the crisis: An assessment, *Management Science* 69, 7328–7345.

Easley, David, Nicholas M Kiefer, Maureen O'Hara, and Joseph B Paperman, 1996, Liquidity, information, and infrequently traded stocks, *The Journal of Finance* 51, 1405–1436.

Easley, David, Marcos M López de Prado, and Maureen O'Hara, 2011, The microstructure of the flash crash: Flow toxicity, liquidity crashes and the probability of informed trading, *Journal of Portfolio Management* 37, 118–128.

——— , 2012, Flow toxicity and liquidity in a high-frequency world, *The Review of Financial Studies* 25, 1457–1493.

Foucault, Thierry, Roman Kozhan, and Wing Wah Tham, 2015, Toxic arbitrage, *The Review of Financial Studies* 28, 1053–1094.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Glosten, Lawrence R, and Lawrence E Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123–142.

Glosten, Lawrence R, and Paul R Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Grossman, Sanford J, and Merton H Miller, 1988, Liquidity and market structure, *The Journal of Finance* 43, 617–633.

Hagströmer, Björn, Lars Nordén, and Dong Zhang, 2014, How aggressive are high-frequency traders?, *Financial Review* 49, 395–419.

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

——— , 1991b, The summary informativeness of stock trades: An econometric analysis, *The Review of Financial Studies* 4, 571–595.

——— , 2010, The best bid and offer: A short note on programs and practices, Working paper New York University.

——— , and Gideon Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646–679.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Hendershott, Terrence, Charles M Jones, and Albert J Menkveld, 2011, Does algorithmic trading improve liquidity?, *The Journal of Finance* 66, 1–33.

Hsieh, David A, and Robert C Merton, 1990, Ldc debt crisis, monetary policy, and financial stability, *The Quarterly Journal of Economics* 105, 445–470.

Huang, Jennifer, and Jiang Wang, 2009, Liquidity provision in limit order markets: The benefits of being patient, *The Review of Financial Studies* 22, 767–793.

——— , 2010, Market liquidity, asset prices, and welfare, *Journal of Financial Economics* 95, 107–127.

Jylhä, Petri, 2018, Margin requirements and the security market line, *The Journal of Finance* 73, 1281–1321.

Kirilenko, Andrei, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2017, The flash crash: High-frequency trading in an electronic market, *The Journal of Finance* 72, 967–998.

Kyle, Albert S, 1985, Continuous auctions and insider trading, *Econometrica* pp. 1315–1335.

——— , and Anna A Obizhaeva, 2016, Market microstructure invariance: Empirical hypotheses, *Econometrica* 84, 1345–1404.

Lazarus, Eben, Daniel J Lewis, James H Stock, and Mark W Watson, 2018, The making of hawks and doves, *Journal of Monetary Economics* 96, 129–143.

Lee, Charles, and Mark J Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Menkveld, Albert J., 2016, The economics of high-frequency trading: Taking stock, *Annual Review of Financial Economics* 8, 1–24.

Moore, Thomas G, 1966, Stock market margin requirements, *Journal of Political Economy* 74, 158–167.

Nagel, Stefan, 2012, Evaporating liquidity, *Review of Financial Studies* 25, 2005–2039.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Odders-White, Elizabeth R, and Mark J Ready, 2008, The probability and magnitude of information events, *Journal of Financial Economics* 87, 227–248.

O'Hara, Maureen, 2015, High frequency market microstructure, *Journal of Financial Economics* 116, 257–270.

Patton, Andrew J, 2009, Correction of the standard errors of regression coefficient estimates when the regressor is measured with error, *Economics Letters* 102, 16–19.

Schwert, G William, 1989, Why does stock market volatility change over time?, *The Journal of Finance* 44, 1115–1153.

van Kervel, Vincent, and Albert J Menkveld, 2018, High-frequency trading around large institutional orders, *Journal of Finance* 73, 2105–2161.

Vayanos, Dimitri, and Jiang Wang, 2013, Market liquidity - theory and empirical evidence, *Handbook of the Economics of Finance* 2, 1289–1361.

V. Figures

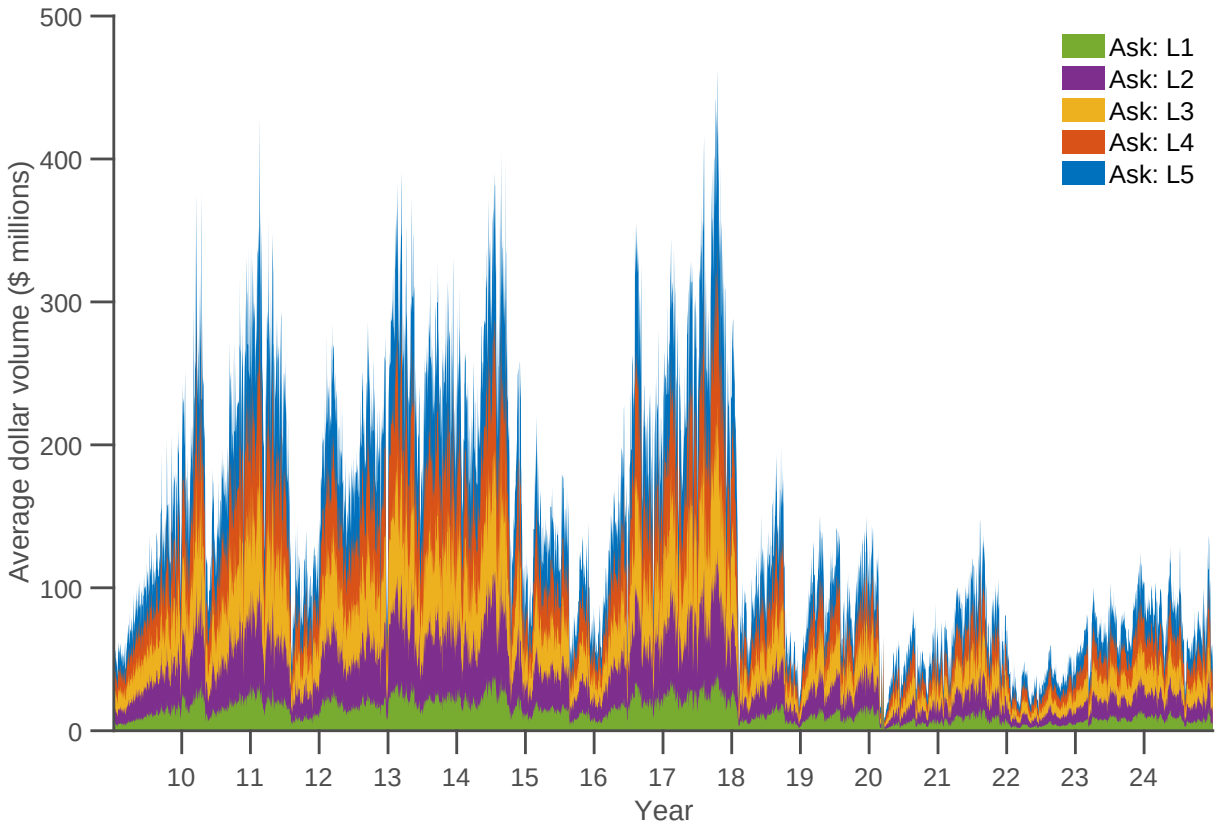
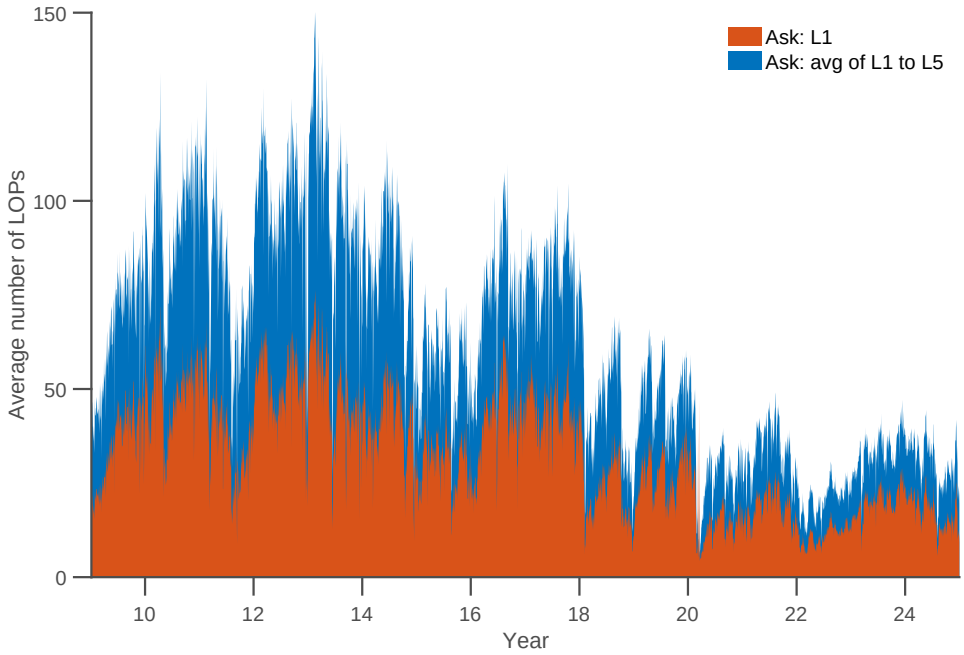
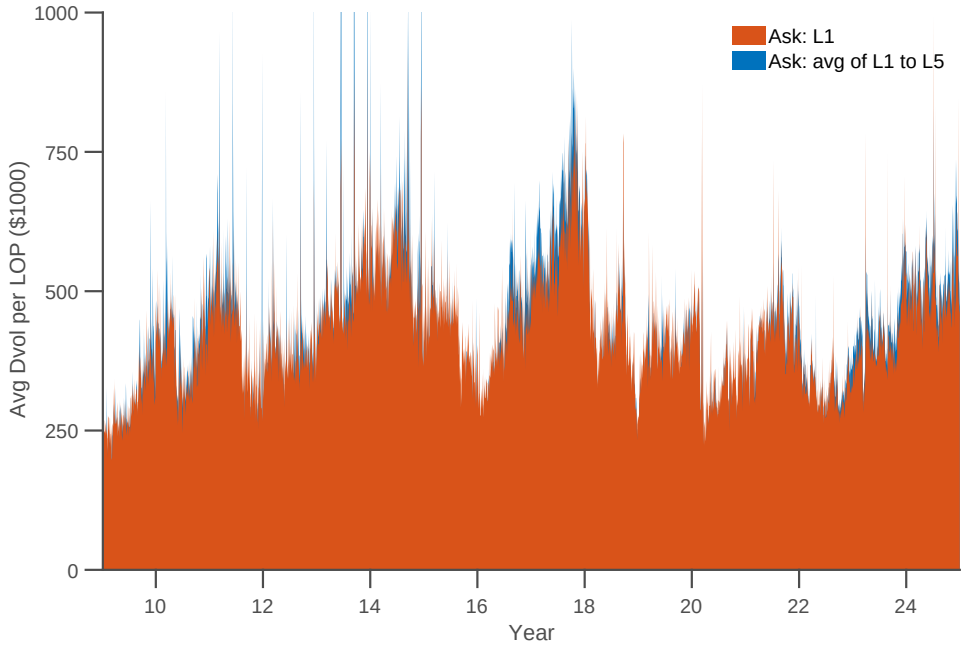


Figure 1. Time series of market depth

The figure plots the time series of market depth measured as dollar volume at level 1 to 5 of the limit order book for the S&P 500 E-mini futures contract. The buy and sell side are almost identical. The daily dollar volume is measured as averages of 5-minute samples throughout the trading day. The sample period is January 2009 to December 2024.



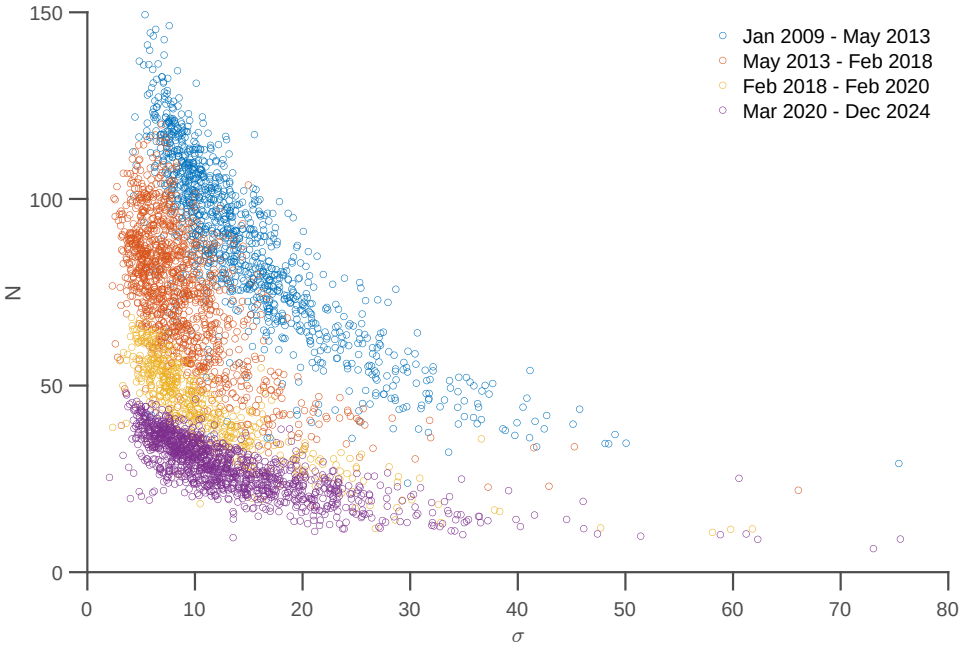
(a) Number of limit order providers (LOPs)



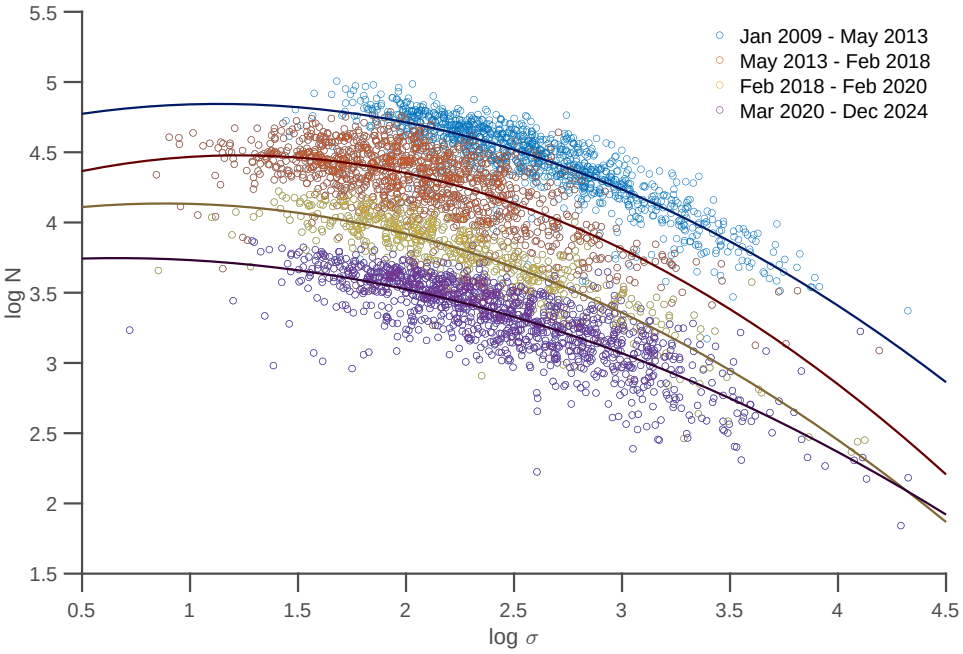
(b) Depth per LOP

Figure 2. Time series of limit order providers and depth per LOP

Figure (a) plots the time series for the number of limit order providers (LOPs) at level 1 of the order book and for the average number of LOPs at level 1 to 5. Figure (b) plots the time series of the average daily dollar volume offered per LOP. The daily LOP number and dollar volume per LOP are measured as averages of 5-minute samples throughout the trading day. The sample period is January 2009 to December 2024.



(a) N versus σ



(b) $\log N$ versus $\log \sigma$

Figure 3. Scatter plot of liquidity providers and volatility

The figure is a scatter plot of the relationship between the average daily number of liquidity providers N (computed as the average number of market participants posting limit orders across level 1 to 5 of the limit order book), and volatility σ (computed as the annualized daily realized volatility from 5-minute log returns) for the S&P 500 E-mini futures contract. Panel (a) plots N versus σ in levels. Panel (b) plots $\log N$ versus $\log \sigma$, where the quadratic fits show that $\log N$ is hump-shaped in $\log \sigma$, becoming linearly decreasing to the left of the turning point. The full sample period is January 2009 to December 2024, split into 4 sub-periods based on the downward shifts of the σ–N curve.

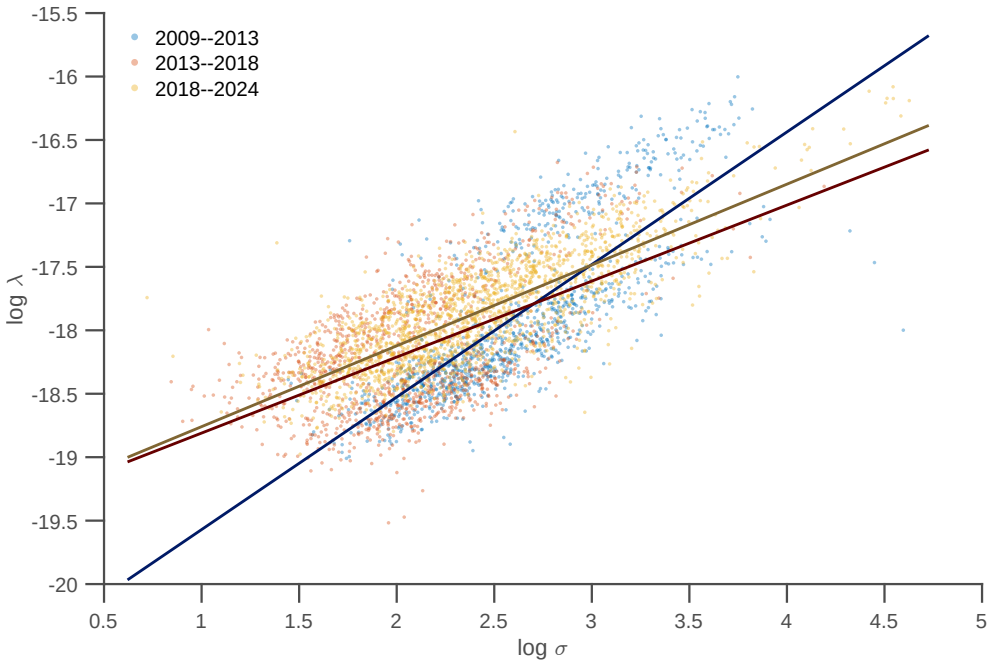


Figure 4. Scatter plot of illiquidity and volatility in logs by sub-period

The figure is a scatter plot of the relationship between $\log \lambda$ (the log of daily Kyle's λ, averaged across U.S. trading hours) and $\log \sigma$ (the log of annualized daily realized volatility from 5-minute log returns) for the S&P 500 E-mini futures contract. The model (Prediction 2) predicts $\log \lambda = \log \sigma - \log \sigma_z$, i.e., a unit slope with intercept $-\log \sigma_z$. The data are split into 3 sub-periods: 2009–2013 (blue), 2013–2018 (red), and 2018–2024 (gold), with separate OLS regression lines. See Table II for sub-period regression statistics.

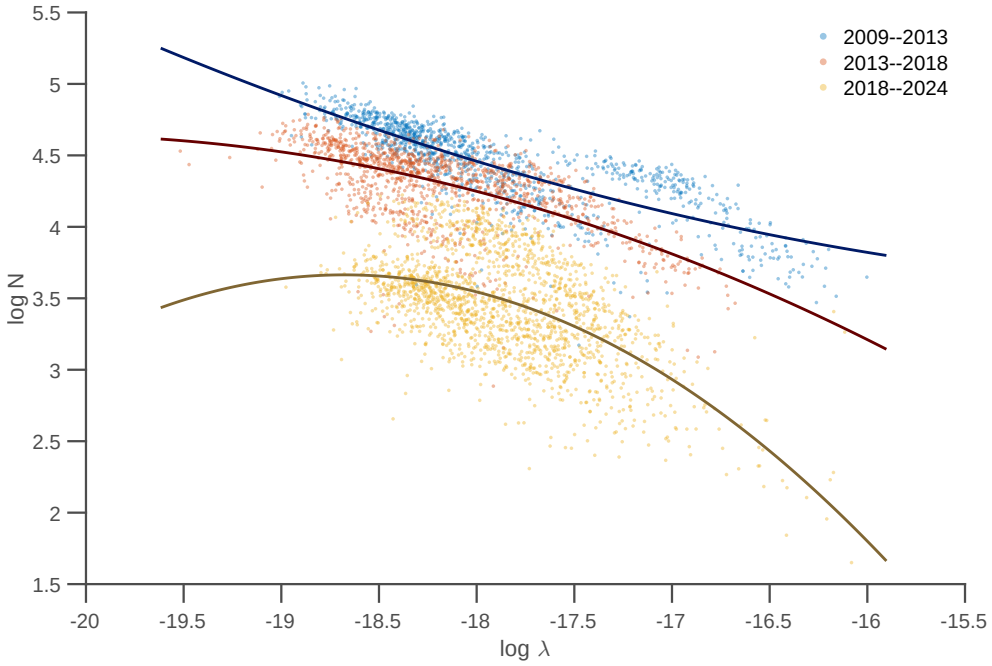


Figure 5. Scatter plot of liquidity providers and illiquidity in logs by sub-period The figure is a scatter plot of the relationship between $\log N$ (the log of the average daily number of liquidity providers across level 1 to 5 of the limit order book) and $\log \lambda$ (the log of daily Kyle's λ, averaged across U.S. trading hours) for the S&P 500 E-mini futures contract. The model (Prediction 3) predicts a hump-shaped relationship. The data are split into 3 sub-periods: 2009–2013 (blue), 2013–2018 (red), and 2018–2024 (gold), with separate quadratic fits $\log N = a + b_1 \log \lambda + b_2 (\log \lambda)^2$. See Table III for sub-period regression statistics.

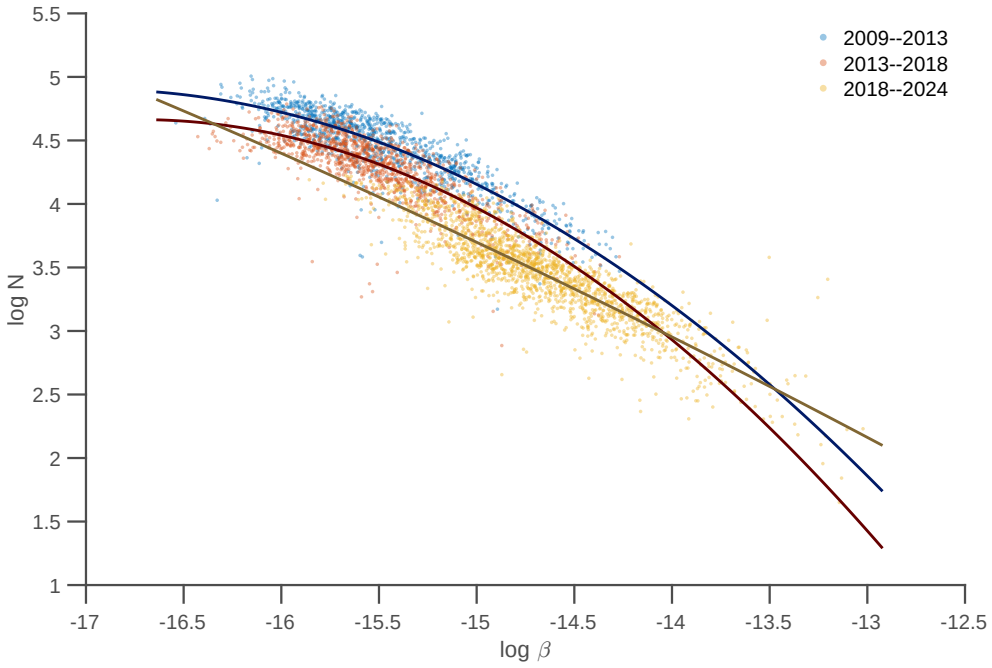


Figure 6. Scatter plot of liquidity providers and price impact in logs by sub-period The figure is a scatter plot of the relationship between $\log N$ (the log of the average daily number of liquidity providers across level 1 to 5 of the limit order book) and $\log \hat{\beta}$ (the log of the estimated daily price impact), for the S&P 500 E-mini futures contract. Daily price impact $\hat{\beta}_t$ is computed as the average across U.S. trading hours of hourly OLS regressions of 1-minute log returns on 1-minute order flow. The model (Prediction 4) predicts a hump-shaped relationship. The data are split into 3 sub-periods: 2009–2013 (blue), 2013–2018 (red), and 2018–2024 (gold), with separate quadratic fits. See Table IV for sub-period regression statistics.

VI. Tables

Hour	09-10	10-11	11-12	12-13	13-14	14-15	15-16	CTC	CTO	OTC
intercept	5.01	5.08	5.22	5.49	5.10	5.61	5.47	4.98	4.91	5.23
se	0.61	0.36	0.27	0.20	0.34	0.26	0.27	0.44	0.63	0.44
$\log \sigma$	0.50	0.58	0.46	0.17	0.38	0.04	0.18	0.19	−0.08	0.39
se	0.40	0.24	0.19	0.15	0.24	0.18	0.18	0.32	0.52	0.29
$(\log \sigma)^2$	−0.16	−0.16	−0.15	−0.11	−0.13	−0.07	−0.10	−0.14	−0.13	−0.14
se	0.06	0.04	0.03	0.03	0.04	0.03	0.03	0.06	0.11	0.05
R^2	63.88	64.05	63.00	64.35	45.21	66.15	71.04	72.77	64.91	65.50

(a) 2009 - 2013

Hour	09-10	10-11	11-12	12-13	13-14	14-15	15-16	CTC	CTO	OTC
intercept	3.52	3.81	4.49	4.89	4.69	5.10	5.20	4.17	4.04	4.29
se	0.46	0.31	0.26	0.19	0.29	0.13	0.21	0.32	0.28	0.36
$\log \sigma$	1.30	1.27	0.82	0.46	0.55	0.23	0.22	0.58	0.27	0.93
se	0.34	0.25	0.23	0.18	0.25	0.11	0.17	0.29	0.28	0.29
$(\log \sigma)^2$	−0.31	−0.30	−0.24	−0.19	−0.19	−0.13	−0.13	−0.23	−0.18	−0.27
se	0.06	0.05	0.05	0.04	0.05	0.02	0.03	0.06	0.07	0.06
R^2	42.55	42.37	47.99	49.26	31.25	55.88	59.95	49.55	46.77	48.76

(b) 2013 - 2018

Hour	09-10	10-11	11-12	12-13	13-14	14-15	15-16	CTC	CTO	OTC
intercept	4.95	3.95	4.00	4.34	4.30	4.41	4.74	4.50	4.21	4.29
se	0.46	0.38	0.35	0.27	0.24	0.30	0.35	0.26	0.20	0.36
$\log \sigma$	−0.16	0.75	0.69	0.39	0.37	0.33	0.14	−0.17	−0.36	0.48
se	0.30	0.25	0.25	0.21	0.19	0.24	0.23	0.20	0.19	0.24
$(\log \sigma)^2$	−0.07	−0.22	−0.22	−0.17	−0.16	−0.15	−0.12	−0.08	−0.04	−0.18
se	0.05	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.04	0.04
R^2	66.72	70.72	67.35	67.52	56.22	71.55	75.53	71.24	69.17	71.07

(c) 2018 - 2024

Table I. Volatility and the number of liquidity providers

This table shows the coefficients for regressions of the type

$$\log N_{t,h} = \alpha + \beta_1 \log \sigma_{t,h} + \beta_2 (\log \sigma_{t,h})^2 + \epsilon_t,$$

where $N_{t,h}$ is the number of limit order providers during trading period h on day t, and $\sigma_{t,h}$ is the annualized realized volatility computed from 1-minute log returns. Standard errors are estimated using a block bootstrap with 10,000 draws, where the optimal block size is chosen following Patton (2009). Each panel reports one sub-period: 2009–2013, 2013–2018, and 2018–2024.

	a	b	SE(b)	$t(b=1)$	$R^2(\%)$	N
2009–2013	-20.615	1.045	0.079	0.56	59.7	1089
2013–2018	-19.409	0.599	0.071	-5.64	30.5	1169
2018–2024	-19.396	0.637	0.026	-13.82	63.4	1739

Table II. Illiquidity and volatility: sub-period regressions

This table reports OLS regressions of $\log \lambda$ on $\log \sigma$ for the S&P 500 E-mini futures contract, estimated separately for each sub-period. The regression is

$$\log \lambda_t = a + b \cdot \log \sigma_t + \epsilon_t,$$

where λ_t is the daily average of hourly Kyle's λ during U.S. trading hours and σ_t is the annualized daily realized volatility from 5-minute log returns. The model (Prediction 2) predicts $b = 1$ and intercept $a = -\log \sigma_z$. The column $t(b = 1)$ reports the t-statistic for the null hypothesis $H_0 : b = 1$ (not $b = 0$). Standard errors use the Newey and West (1987) correction.

	a	b_1	b_2	SE(b_2)	$t(b_2)$	$\log \lambda^*$	$R^2(\%)$
2009–2013	12.295	1.284	0.047	0.032	1.46	-13.62	67.6
2013–2018	-28.379	-3.268	-0.081	0.044	-1.85	-20.21	41.9
2018–2024	-87.031	-9.712	-0.260	0.052	-4.98	-18.68	35.0

Table III. LP number and illiquidity: sub-period quadratic regressions

This table reports quadratic regressions of $\log N$ on $\log \lambda$ for the S&P 500 E-mini futures contract, estimated separately for each sub-period. The regression is

$$\log N_t = a + b_1 \cdot \log \lambda_t + b_2 \cdot (\log \lambda_t)^2 + \epsilon_t.$$

The model (Prediction 3) predicts a hump-shaped relationship, i.e., $b_2 < 0$. The turning point $\log \lambda^* = -b_1/(2b_2)$ denotes the level of illiquidity at which the LP number peaks. Standard errors use the Newey and West (1987) correction.

	a	b_1	b_2	SE(b_2)	$t(b_2)$	$\log \hat{\beta}^*$	$R^2(\%)$	N
2009–2013	-50.778	-6.564	-0.193	0.040	-4.86	-16.97	76.5	1097
2013–2018	-60.540	-7.796	-0.233	0.046	-5.03	-16.73	68.8	1180
2018–2024	-12.227	-1.400	-0.023	0.023	-0.99	-31.00	79.2	1739

Table IV. LP number and price impact: sub-period quadratic regressions

This table reports quadratic regressions of $\log N$ on $\log \hat{\beta}$ for the S&P 500 E-mini futures contract, estimated separately for each sub-period. The regression is

$$\log N_t = a + b_1 \cdot \log \hat{\beta}_t + b_2 \cdot (\log \hat{\beta}_t)^2 + \epsilon_t.$$

The daily price impact $\hat{\beta}_t$ is estimated as the average across U.S. trading hours of hourly OLS regressions of 1-minute log returns on 1-minute order flow. The model (Prediction 4) predicts a hump-shaped relationship, i.e., $b_2 < 0$. The turning point $\log \hat{\beta}^*$ denotes the price impact level at which the LP number peaks. Standard errors use the Newey and West (1987) correction.

	(i)	(ii)	(iii)	(iv)
$\Delta\kappa_t$	-0.193			-0.128
	(0.043)			(0.032)
ΔCR_t		0.183		0.126
		(0.037)		(0.030)
ΔCDS_t			-0.144	-0.036
			(0.036)	(0.028)
$R^2(\%)$	3.02	3.47	2.13	4.92

(a) ΔN_t as response

	(i)	(ii)	(iii)	(iv)
$\Delta\kappa_t$	0.204			0.144
	(0.033)			(0.031)
ΔCR_t		-0.183		-0.126
		(0.019)		(0.021)
ΔCDS_t			0.138	0.025
			(0.024)	(0.022)
$R^2(\%)$	6.68	6.86	3.91	10.17

(b) $\Delta\sigma_t$ as response

Table V. LP participation and constraint proxies: addressing endogeneity

This table shows regression coefficients for the daily number of limit order providers (LOPs), N_t, and volatility, σ_t, regressed on daily LP constraint proxies, D_t^i, for the S&P 500 E-mini futures contract. The constraint proxies are plausibly exogenous to the contemporaneous LP participation decision, providing reduced-form evidence that LP entry responds to observable variation in constraints. The regressions are in differences:

$$\Delta N_t = \alpha + \beta \Delta D_t^i + \epsilon_t \qquad \text{and} \qquad \Delta\sigma_t = \alpha + \beta \Delta D_t^i + \epsilon_t.$$

We proxy for LP constraints with (i) the E-mini margin percentage, κ_t, which reflects exchange-level requirements set by the CME, (ii) the capital ratio of NYSE primary dealers, CR_t, from He, Kelly, and Manela (2017), which captures aggregate financial sector health, and (iii) the average credit default swap rate of G14 dealers, CDS_t, which reflects dealer credit risk. Column (iv) includes all three proxies in a multiple regression. All variables are in differences and z-scored. Standard errors are estimated using a Newey and West (1987) correction with lags selected following Lazarus, Lewis, Stock, and Watson (2018). The sample period is January 2009 to December 2024.

Exchange-Traded Liquidity

Online Appendix

Section A.1 contains proofs of the model results. Section A.2 provides an alternative model and shows that the qualitative predictions of the model are the same as of the one in Section I. Section A.3 describes our algorithm for classifying limit order book changes into limit orders, market orders, and cancellations. Section A.4 describes the trader types defined by the CFTC and their positions in the E-mini contract. Section A.5 gives an overview of the most common measures of liquidity. Section A.6 collects supplementary figures and Section A.7 collects supplementary tables.

A.1. Proofs and derivations

A.2. Alternative model

There are three dates, $t = 0$, 1, and 2. The financial market consists of a riskless and a risky asset that can be traded at date 1 and pay off at date 2. The riskless asset is in elastic supply and its riskfree net return is normalized to zero. The risky asset, whose supply is normalized to zero for simplicity, pays off D units, where D has mean $\overline{D}$ and variance σ^2. Using the riskless asset as the numeraire, we denote by p the risky asset's price at date 1.

The economy is populated by two types of agents, end users and liquidity providers (LPs). End users, in exogenously given mass n, have exponential utility with the coefficient of absolute risk aversion α. For convenience, we express their utility as certainty-equivalent wealth, i.e., as

$$-\frac{1}{\alpha}\log \mathrm{E}\left[\exp\left\{-\alpha W_2^e\right\}\right], \tag{A.1}$$

where W_2^e is the terminal (date-2) wealth. In turn, their terminal wealth consists of the starting wealth W^e, capital gains from trading the risky asset, and an endowment $z\left(D-\overline{D}\right)$ of the consumption good at date 2 that covaries with the payoff of the risky asset. Formally,

$$W_2^e = W^e + x_1^e\left(D-p\right) + z\left(D-\overline{D}\right), \tag{A.2}$$

where x_1^e denotes their position in the risky asset. We assume that the variable z is Gaussian with mean zero and variance σ_z^2, and is independent of D. While the endowment is received at date 2, agents learn z at date 1. Since the endowment is correlated with D, it generates a hedging demand. When, for example, $z > 0$, the endowment exposes agents to the risk that D will be low, and agents hedge against that risk by selling the risky asset. We assume that the endowment is perfectly correlated with D for simplicity; what matters for our analysis is that the correlation is non-zero and hence generates trade.

LPs are assumed to be risk neutral and do not receive any endowment shocks, but are subject to a constraint on the position they can take. Formally, they maximize their expected date-2 wealth,

$$\mathrm{E}\left[W_2^d\right], \tag{A.3}$$

subject to the budget constraint

$$W_2^d = W^d + x_1^d\left(D-p\right) \tag{A.4}$$

and the margin constraint

$$\kappa \left|x_1^d\right| \leq W^d, \tag{A.5}$$

where W^d denotes their starting wealth, x_1^d is their position in the risky asset, and $\kappa \geq 0$ is the capital requirement associated with trading the risky asset. The margin constraint (A.5) states that in order to undertake a trade of size $|x_1^d|$ LPs must commit $\kappa \left|x_1^d\right|$ of their scarce capital.

To simplify the notation, we define $w \equiv W^d/\kappa$ as the balance sheet space LPs can allocate to the risky asset. This implies that (A.5) is equivalent to

$$\left|x_1^d\right| \leq w. \tag{A.6}$$

In equilibrium, end users initiate trades with LPs to share risk. Because they can be thought of as consuming market liquidity, we also refer to them as liquidity demanders. Moreover, we refer to z as the liquidity shock.

In a frictionless economy, all agents would be present in all markets periods. Thus, any seller of an asset, for example, can have immediate access to the entire population of buyers. In practice, however, agents face costs of market participation. Such costs include buying trading infrastructure or membership of a financial exchange, having capital available on short notice, monitoring market movements, etc. To obtain a limited liquidity supply, we assume costly participation of LPs in the form of incurring a cost c to enter the market and trade at date 1. Consistent with the notion that participation is an ex-ante decision, we assume that agents must decide whether or not to incur c at date 0, i.e., before observing the price at date 1.[18] We denote the mass of available LPs by $N_{max} \in (0, \infty]$, also allowing for an infinite pool of them, and by N the mass of LPs actually participating in the market and providing liquidity at date 1.

A. Equilibrium

In this section we first solve for the equilibrium of the model. We go backwards from date 1 to date 0. We first compute the equilibrium price and quantities at date 1 given the mass of market participants, and then we study the entry decision of LPs at date 0.

At date 1, end-users choose holding x_1^e of the risky asset to maximize the expectation of (A.1). Substituting (A.2) into (A.1), after some algebra we obtain that the optimization problem of end-users is equivalent to

$$\max_{x_1^e} W^e + x_1^e \left(\overline{D} - p\right) - \frac{1}{2}\alpha\sigma^2 \left(x_1^e + z\right)^2. \tag{A.7}$$

From here, optimal demand of liquidity demanders is given by

$$x_1^e = \frac{\overline{D} - p}{\alpha\sigma^2} - z. \tag{A.8}$$

It has two components: a downward-sloping demand curve, according to which they buy the asset as long as the price is below the fundamental value and (short) sell otherwise, and a term influenced by the liquidity shock z. When, for example, z is positive and large, liquidity demanders are willing to sell because their endowment is positively correlated with the asset.

A liquidity supplier chooses holdings x_1^d of the risky asset to maximize (A.3). Substituting in

[18] If the decision can be made contingent on the price at date 1, then c is a fixed transaction cost rather than a participation cost. We consider transaction costs as an alternative in the appendix.

(A.4) and taking the constraint (A.5) into account, after some algebra we obtain

$$x_1^d = \begin{cases} -w & \text{if } p > \overline{D} \\ \in [-w, w] & \text{if } p = \overline{D} \\ w & \text{if } p < \overline{D}. \end{cases} \tag{A.9}$$

LPs are willing to buy the risky asset as long as it trades below its expected payoff $\overline{D}$ and are willing to sell otherwise. Since they are risk neutral, they trade up to the maximum amount allowed by their leverage constraints.

Market clearing requires that aggregate demand equals supply, i.e.,

$$N \cdot sgn\left(\overline{D} - p\right) \cdot w + n \cdot \left(\frac{\overline{D} - p}{\alpha\sigma^2} - z\right) = 0,$$

which implies

$$p = \begin{cases} \overline{D} - \alpha\sigma^2\left(z + \frac{N}{n}w\right) & \text{if } z < -\frac{N}{n}w \\ \overline{D} & \text{if } |z| \leq \frac{N}{n}w \\ \overline{D} - \alpha\sigma^2\left(z - \frac{N}{n}w\right) & \text{if } z > \frac{N}{n}w. \end{cases} \tag{A.10}$$

Naturally, the finite mass of LPs and their position limits implies that they can fully absorb the liquidity shock z as long as the shock is small. In this case, LPs drive the market price back to its expectation, $\overline{D}$. In contrast, when the liquidity shock is large, end users can only partially share risk with LPs and require a premium for bearing risk.

Next we consider LPs' liquidity supply decision. Due to the risk neutrality assumption, their aggregate position is either $\frac{N}{n}w$ or $-\frac{N}{n}w$, and, as long as they decide to enter the market, their expected interim utility (after the liquidity shock realization but before the asset payoff) is

$$\mathcal{U}_1^d(z) = \begin{cases} W^d - c - w\alpha\sigma^2\left(z + \frac{N}{n}w\right) & \text{if } z < -\frac{N}{n}w \\ W^d - c & \text{if } |z| \leq \frac{N}{n}w \\ W^d - c + w\alpha\sigma^2\left(z - \frac{N}{n}w\right) & \text{if } z > \frac{N}{n}w, \end{cases} \tag{A.11}$$

which consist of the starting wealth, the participation cost c, and the trading gain calculated from (A.9) and (A.10). After some algebra, we obtain that the ex ante expected utility of LPs who enter the market is

$$\mathcal{U}_0^d = W^d - c + 2\alpha\sigma^2\sigma_z w\Psi\left(-\frac{1}{\sigma_z}\frac{N}{n}w\right), \tag{A.12}$$

where

$$\Psi(x) \equiv \phi(x) + x \cdot \Phi(x) \tag{A.13}$$

is a strictly increasing convex function, and ϕ and Φ denote the pdf and cdf of the standard normal distribution.

Equations (A.11) and (A.12) highlight the two effects that higher LP balance sheet space has on realized and expected profits. Higher wealth or lower margins allow risk neutral LPs to take larger positions, which in turn brings the price closer to the expected asset payoff and lowers the realized and expected dollar return. As a result, the expected profit is a hump-shaped function of w. Using (A.12), we can show that higher balance sheet space allows for higher expected profits as long as

$$0 \leq w \leq \bar{w}, \tag{A.14}$$

where $\overline{w} > 0$ denotes the only root of $0 = \phi\left(-\frac{N}{n}\frac{\overline{w}}{\sigma_z}\right) + 2\left(-\frac{N}{n}\frac{\overline{w}}{\sigma_z}\right)\Phi\left(-\frac{N}{n}\frac{\overline{w}}{\sigma_z}\right)$, assuming $N/(n\sigma_z) > 0$. From now on we make the assumption that we are in this intuitive range.

LPs enter the market as long as their expected utility is at least as high as the utility from staying out:

$$\mathcal{U}_0^d \geq \overline{\mathcal{U}}_0^d \equiv W^d. \tag{A.15}$$

Since we require $N \geq 0$, (A.12) and (A.15) imply the following result:

Proposition A.1. (Balance sheet space and LP number) *Let* $\overline{c}(w) \equiv \sqrt{2/\pi}\ \alpha\sigma^2\sigma_z w$. *If* $c \geq \overline{c}(w)$, *we have* $N = 0$. *If* $0 < c < \overline{c}(w)$, *LPs enter the market as long as their equilibrium mass reaches*

$$N = -n\sigma_z \frac{1}{w}\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w}\right) > 0. \tag{A.16}$$

If (A.14) also holds, the equilibrium LP number N *is increasing in LPs' wealth and decreasing in the margin,* $\partial N/\partial w > 0$.

Proposition A.1 states that if the entry cost is too high compared to LPs' available balance sheet space $w \equiv W^d/\kappa$, they will stay out of the market, but if the entry cost is moderate, we obtain a finite number of LPs deciding to provide liquidity in equilibrium. In turn, the equilibrium number of LPs is increasing in LP balance sheet space.

Our model provides testable predictions regarding the relationship between return moments and LP numbers. First, we focus on the level of returns. Let us assume that at date 0 the price is the unconditional expectation of the asset, $\overline{D}$–this would also be in line with the assumption that LPs are risk neutral and, absent any constraints, any infinitesimal measure of them would drive the price back to $\overline{D}$.

From (A.10), let us refer to *realized returns* calculated as those between dates 0 and 1 and given by $R_1 = p - \overline{D}$, and *expected returns* as those between dates 1 and 2, conditional on the liquidity shock: $\mathrm{E}[R_2|z] = \mathrm{E}[R_2|p] = \overline{D} - p = -R_1$. Since dealers are assumed to be risk neutral and they are subject to the balance sheet constraint (A.6), it is imminent that the unconditional expected dollar returns satisfy $\mathrm{E}[R_2] = -\mathrm{E}[R_1] = 0$. However, conditional on dealers' equilibrium positions, returns must be high enough to compensate them for the cost of liquidity provision, c. For example, as long as there is a large positive liquidity shock ($z > (N/n)\cdot w$) and end users dump the asset which dealers are expected to hold, this pushes the date-1 price down, leading to positive expected future returns, both in terms of what an econometrician observes, $\mathrm{E}[R_2|z]$, and what dealers' expect on their long positions. In contrast, as long as there is a large negative liquidity shock ($z < -(N/n)\cdot w$) and end users buy the asset which dealers are expected to supply, this pushes the date-1 price up, leading to lower (in fact, negative) future expected returns $\mathrm{E}[R_2|z]$, but again positive expected returns for dealers who are supplying the asset. Formally, we obtain the following results:

Proposition A.2. (Balance sheet space and returns) *Contemporaneous and expected future returns satisfy* $E[|R_2|] = E[|R_1|] = \frac{c}{w}$, *with* $sgn(E[R_2|z]) = -sgn(R_1) = sgn(z)$. *Therefore, (i) higher balance sheet space lowers absolute realized returns and lowers absolute expected future returns, and (ii) the variable constructed as an interaction of balance sheet space and the sign of dealers' position has a positive relationship with realized returns and negatively predicts expected future returns:* $dE[R_1]/d(w \cdot sgn(z)) > 0$ *and* $dE[R_2]/d(w \cdot sgn(z)) < 0$.

Proposition A.1 and A.2 together imply that dealer numbers have the same relationship to returns as the balance sheet space:

Corollary A.1. (Dealer number and returns) *Assuming the variation in the equilibrium dealer number is driven by variation in balance sheet space, fewer dealers leads to larger price reversals. Specifically, there is a positive relationship between contemporaneous returns and dealer number multiplied by the sign of the dealers' positions, and a negative relationship between expected future returns and dealer number multiplied by the sign of dealers' positions:* $dE[R_1]/d(N \cdot sgn(z)) > 0$ *and* $dE[R_2]/d(N \cdot sgn(z)) < 0$.

Next, we combine Proposition A.1 with (A.10) to express the equilibrium price volatility at date 1 in terms of model primitives:

Proposition A.3. (Balance sheet space and volatility) *When* $c \geq \bar{c}(w)$, *price volatility is* $Var[p] = (\alpha\sigma^2)^2 \sigma_z^2$. *When* $0 < c < \bar{c}(w)$,

$$Var[p] = 2(\alpha\sigma^2)^2 \sigma_z^2 \Gamma\left(\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w}\right)\right), \tag{A.17}$$

where $\Gamma(x) \equiv \Phi(x) + x \cdot \Psi(x)$ *is a strictly increasing function. As a result,* $\partial Var[p]/\partial w < 0$.

Then, Propositions A.1 and A.3 allow us to obtain the following result regarding the relationship of price volatility and equilibrium dealer number as a result of time-variation in w while keeping other model parameters constant:

Corollary A.2. (Dealer number and volatility) *Assume that* $0 < c < \bar{c}(w)$. *Then we have* $dN/dVar[p] < 0$.

Suppose we allow w to be time varying while keeping other model parameters constant. As a result, Corollary A.2 implies a negative relationship between return volatility and the equilibrium number of liquidity providers, and the mechanism is through the impact of margin constraints dealers are subject to: more balance sheet space (lower dealer wealth and/or higher margins) lead to a lower number of liquidity providers and higher volatility.

Next, we examine how the relationship between volatility, the number of liquidity providers, and available balance sheet capital changes if the participation cost increases:

Proposition A.4. *Assume that* $0 < c < \bar{c}(w)$ *so that* $N > 0$. *An increase in* c *while keeping* c/w *constant lowers* N *but does not affect* $Var[p]$.

A straightforward implication of Proposition A.4 is the following:

Corollary A.3. ($N-$Var $[p]$ curve shifts) *When participation gets more costly and* c *increases, the* $N - Var[p]$ *curve shifts "inwards", and the same level of* $Var[p]$ *corresponds to a lower equilibrium LP number* N.

We can also relate the equilibrium LP number to other standard microstructure liquidity measures. For example, we express Kyle's lambda in our model measured as the equilibrium price impact of liquidity shocks. From (A.10), it is imminent that

$$\frac{\partial p}{\partial z} = \begin{cases} -\alpha\sigma^2 & \text{if } z < -\frac{N}{n}w \\ 0 & \text{if } |z| \leq \frac{N}{n}w \\ -\alpha\sigma^2 & \text{if } z > \frac{N}{n}w. \end{cases} \tag{A.18}$$

We define the (average) Kyle's lambda in the model as $\lambda \equiv \mathrm{E}[|\partial p/\partial z|]$. Combining the definition with (A.16), after some algebra we obtain the following result:

Proposition A.5. (LP number and Kyle's lambda) *The average Kyle's lambda in our model is given by*

$$\lambda \equiv E\left[\left|\frac{\partial p}{\partial z}\right|\right] = 2\alpha\sigma^2\left[1-\Phi\left(\frac{N}{n}\frac{w}{\sigma_z}\right)\right]. \tag{A.19}$$

As a result, higher available balance sheet space decreases the average price impact, $\partial\lambda/\partial w < 0$. As long as (A.14) also holds, Kyle's lambda and the equilibrium LP number N have a negative relationship, $d\lambda/dN < 0$.

B. Proofs and derivations

We start with some useful results.

Lemma A.1. *The function*

$$\Psi(x) \equiv \phi(x) + x\Phi(x) \tag{A.20}$$

has the following three sets of properties:

(i) $\Psi(x) > 0\ \forall x \in R$, *and its limits are* $\lim_{x\to-\infty}\Psi(x) = 0$ *and* $\lim_{x\to\infty}\Psi(x) = \infty$.

(ii) $\Psi(x)$ *is increasing; its slope satisfies* $0 < \Psi'(x) < 1$ *for all* $x \in R$ *with limits* $\lim_{x\to-\infty}\Psi'(x) = 0$ *and* $\lim_{x\to\infty}\Psi'(x) = 1$, *and* Ψ' *increases monotonically in between.*

(iii) $\Psi(x)$ *is convex:* $\Psi''(x) > 0$ *for all* $x \in R$.

Proof. From (A.20), we have

$$\Psi'(x) = -x\phi(x) + x\phi(x) + \Phi(x) = \Phi(x) > 0$$

and $\Psi''(x) = \phi(x) > 0$. From here (ii) and (iii) are straightforward.

In addition,

$$\lim_{x\to\infty}\Psi(x) = \lim_{x\to\infty}\phi(x) + \lim_{x\to\infty}x\Phi(x) = \infty.$$

At the same time,

$$\begin{aligned}\lim_{x\to-\infty}\Psi(x) &= \lim_{x\to-\infty}\phi(x) + x\Phi(x) = \lim_{x\to-\infty}\frac{\phi(x)\frac{1}{x} + \Phi(x)}{\frac{1}{x}} \\ &= \lim_{x\to-\infty}\frac{-x\phi(x)\frac{1}{x} - \phi(x)\frac{1}{x^2} + \phi(x)}{-\frac{1}{x^2}} = \lim_{x\to-\infty}\phi(x) = 0,\end{aligned}$$

where the second equality is due to l'Hopital's rule. This completes the proof of Lemma A.1. □

Lemma A.2. *For any distribution with pdf f and cdf F we have*

$$\int_k^\infty (z-k)\,f(z)\,dz = [1-F(k)]\left(\int_k^\infty z\frac{f(z)}{1-F(k)}dz - k\right) = [1-F(k)]\cdot\left(E[z|z\geq k] - k\right),$$

Proof. Follows from simple algebra. □

Lemma A.3. *For any distribution with pdf f and cdf F we have*

$$\int_k^\infty (z-k)^2 f(z)\,dz = [1-F(k)]\left(\int_k^\infty (z-k)^2 \frac{f(z)}{1-F(k)}dz\right) = [1-F(k)]\cdot E\left[(z-k)^2 \,|z>k\right]$$

Proof. Follows from simple algebra. □

We are now ready to prove the propositions of the model.

Proof of Proposition A.1. From (A.11), we can write ex ante expected LP utility as

$$\begin{aligned}
\mathcal{U}_0^d &= W^d - c + \int_{-\infty}^{-\frac{N}{n}w_\kappa^d} \left[-w_\kappa^d \alpha\sigma^2\left(z+\frac{N}{n}w_\kappa^d\right)\right]\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\
&\quad + \int_{\frac{N}{n}w_\kappa^d}^{\infty} w_\kappa^d \alpha\sigma^2\left(z-\frac{N}{n}w_\kappa^d\right)\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\
&= W^d - c + 2\alpha\sigma^2 w_\kappa^d \int_{\frac{N}{n}w_\kappa^d}^{\infty}\left(z-\frac{N}{n}w_\kappa^d\right)\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\
&= W^d - c + 2\alpha\sigma^2\sigma_z w_\kappa^d\left[\phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) + \left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\right] \\
&= W^d - c + 2\alpha\sigma^2\sigma_z w_\kappa^d \Psi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right),
\end{aligned} \tag{A.21}$$

where the first equality follows from (A.10), the second equality follows from the symmetric distribution of z, the third equality follows from Lemma A.2, which, applied to a normally distributed random variable, simplifies to

$$\int_k^\infty (z-k) f(z)\,dz = \left[1-\Phi\left(\frac{k}{\sigma_z}\right)\right]\left[0-\sigma_z\frac{-\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}-k\right] = \sigma_z\left[\phi\left(\frac{k}{\sigma_z}\right)+\left(-\frac{k}{\sigma_z}\right)\Phi\left(-\frac{k}{\sigma_z}\right)\right],$$

and the last equality follows from the definition of $\Psi(.)$, (A.13). From (A.21), after some algebra we obtain that (A.15) holds if and only if

$$0 \le N \le -n\sigma_z\frac{1}{w_\kappa^d}\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\right). \tag{A.22}$$

But since $c \ge \bar{c}\left(w_\kappa^d\right) \equiv \sqrt{2/\pi}\alpha\sigma^2\sigma_z w_\kappa^d$ is equivalent to

$$\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d} \ge \frac{1}{\sqrt{2\pi}} = \Psi(0),$$

i.e.,

$$\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\right) \ge 0,$$

we obtain that $c \geq \bar{c}\left(w_\kappa^d\right)$ implies $N = 0$ in equilibrium, whereas $0 < c < \bar{c}\left(w_\kappa^d\right)$ implies (A.16). Differentiating (A.16) with respect to w_κ^d, after some algebra we obtain

$$\begin{aligned}\frac{\partial N}{\partial w_\kappa^d} &= n\sigma_z \frac{1}{\left(w_\kappa^d\right)^2}\left[\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\right) + \frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\frac{1}{\Psi'\left(\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\right)\right)}\right] \\ &= n\sigma_z \frac{1}{\left(w_\kappa^d\right)^2}\frac{\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) + \Psi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)}{\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)},\end{aligned} \tag{A.23}$$

where we used $\left(\Psi^{-1}\right)'(x) = 1/\Psi'\left(\Psi^{-1}(x)\right) = 1/\Phi\left(\Psi^{-1}(x)\right) > 1$. Therefore, $\partial N/\partial w_\kappa^d > 0$ is equivalent to

$$0 < \left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) + \Psi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) = \phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) + 2\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right),$$

which is exactly condition (A.14). □

Proofs of Proposition A.2 and Corollary A.1. Straightforward from risk neutrality and Proposition A.1. □

Proof of Proposition A.3. Combining Proposition A.1 with (A.10) allows us to express the equilibrium price volatility at date 1 in terms of model primitives. In particular, we obtain that $\mathrm{Var}[p] = 0$ when $c = 0$ and $\mathrm{Var}[p] = \left(\alpha\sigma^2\right)^2\sigma_z^2$ when $c \geq \bar{c}\left(w_\kappa^d\right)$. Finally, suppose $0 < c < \bar{c}\left(w_\kappa^d\right)$. Since $z \sim N\left(0, \sigma_z^2\right)$, (A.10) implies that $\mathrm{E}[p] = \overline{D}$. From here, using that the probability density function of z is $\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)$, we can write

$$\begin{aligned}\mathrm{Var}[p] &= \int_{-\infty}^{-\frac{N}{n}w_\kappa^d}\left[-\alpha\sigma^2\left(z + \frac{N}{n}\frac{W_0^d}{\kappa}\right)\right]^2\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz + \int_{\frac{N}{n}w_\kappa^d}^{\infty}\left[-\alpha\sigma^2\left(z - \frac{N}{n}w_\kappa^d\right)\right]^2\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\ &= 2\left(\alpha\sigma^2\right)^2\int_{\frac{N}{n}w_\kappa^d}^{\infty}\left(z - \frac{N}{n}w_\kappa^d\right)^2\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\ &= 2\left(\alpha\sigma^2\right)^2\sigma_z^2\left(\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\left[1 + \left(\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)^2\right] + \left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\right) \\ &= 2\left(\alpha\sigma^2\right)^2\sigma_z^2\left(\Phi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right) + \left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\Psi\left(-\frac{1}{\sigma_z}\frac{N}{n}w_\kappa^d\right)\right) \\ &= 2\left(\alpha\sigma^2\right)^2\sigma_z^2\Gamma\left(\Psi^{-1}\left(\frac{c}{2\alpha\sigma^2\sigma_z}\frac{1}{w_\kappa^d}\right)\right),\end{aligned}$$

where the first equality follows from (A.10), the second equality follows from the symmetric distribution of z, the third equality follows from Lemma A.3, which, applied to a normally distributed

random variable, simplifies to

$$
\begin{aligned}
&\int_{\frac{N}{n}w_\kappa^d}^{\infty}\left(z-\frac{N}{n}w_\kappa^d\right)^2\frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)dz \\
&=\left[1-F\left(k\right)\right]\left(E\left[z^2|z>k\right]-2k\sigma_z\frac{\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}+k^2\right) \\
&=\left[1-F\left(k\right)\right]\left(Var\left[z|z>k\right]+\left(E\left[z|z>k\right]\right)^2-2k\sigma_z\frac{\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}+k^2\right) \\
&=\left[1-\Phi\left(\frac{k}{\sigma_z}\right)\right] \\
&\quad\times\left(\sigma_z^2\left[1-\frac{-\frac{k}{\sigma_z}\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}-\left[\frac{-\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}\right]^2\right]+\left(\sigma_z\frac{\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}\right)^2-2k\sigma_z\frac{\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}+k^2\right) \\
&=\left[1-\Phi\left(\frac{k}{\sigma_z}\right)\right]\left(\sigma_z^2-k\sigma_z\frac{\phi\left(\frac{k}{\sigma_z}\right)}{1-\Phi\left(\frac{k}{\sigma_z}\right)}+k^2\right).
\end{aligned}
$$

The fourth equality follows from the definition of $\Psi(.)$, and in the last step we make use of (A.16) and the definition of $\Gamma(.)$. It is imminent that both $\Gamma(x)$ and $\Psi^{-1}(x)$ are strictly increasing functions, which implies that $\partial \text{Var}[p]/\partial w_\kappa^d < 0$, which completes the proof. □

Proofs of Proposition A.4 and Corollary A.3. From (A.17) it is imminent that as long as c and w_κ^d increase while c/w_κ^d remains constant, $\text{Var}\left[S_1\right]$ remains constant. At the same time N decreases according to (A.16). Since there is a negative relationship between N and $\text{Var}\left[S_1\right]$, this means an increase in c shifts the $N - \text{Var}\left[p\right]$ curve towards the origin. □

A.3. Algorithm for Identifying Order Types

To analyze trading behavior and liquidity provision patterns, we develop a computationally intensive but conceptually simple algorithm to classify changes in the limit order book as limit orders, market orders, or cancellations. This algorithm processes tick-by-tick data, examining every change to the limit order book across all five levels on both the bid and ask sides.

Our raw data consists of snapshots of the limit order book, where each row represents a change to the book. For each level j and side (bid/ask), we observe the price, P_j, the number of contracts, C_j, and the number of LPs posting at that level, N_j. The algorithm works by computing differences between consecutive observations:

Limit Order Identification: When the number of available contracts C_j increases at any level, this unambiguously indicates a new limit order posting. The increase in contracts can only result from a market participant posting additional liquidity at that price level. This classification applies to all levels (1-5) on both bid and ask sides.

Cancellation Identification: When the number of contracts C_j decreases at levels 2-5, this

indicates a limit order cancellation. Since market orders execute at the best available price (level 1), any reduction in depth at lower-priority levels must result from traders canceling their posted orders.

Market Order vs. Cancellation at Level 1: The most challenging classification occurs when contracts decrease at level 1. This could result from either: (i) a market order executing against the posted limit orders, or (ii) a trader canceling their limit order at the top of the book. Our depth-only dataset does not directly distinguish between these cases, but merging our depth data with trade data we are able to separate market orders from cancellations at level 1. This involves computing the number of executed trades in each 5-minute interval and subtracting this from observed decreases in level 1 contracts to approximate cancellations.

Our algorithm processes millions of limit order book changes daily, generating statistics on: (i) the frequency of changes at each level, (ii) the average size of limit orders when posted, and (iii) the distribution of trading activity across different levels of the book. This granular analysis reveals significant time variation in trading behavior and liquidity provision patterns, as discussed in subsequent sections.

The computational intensity of this approach–requiring processing of terabyte-scale datasets–explains why such detailed analysis of order flow has been limited in prior literature. However, the resulting measures provide real-time, forward-looking indicators of market liquidity that complement traditional price-based measures.

A.4. CFTC Data

This section describes the trader types and their market positions in the E-mini market. As per CFTC, the trader types in the futures markets consists of dealers, asset managers, leveraged hedge funds, 'other reportable' and non reportable. Dealers make up the supply side of the market. They seek to have matched books or offset their risk across markets and clients. Dealers of futures contracts consists of large banks (U.S. and non-U.S.) and dealers in securities, swaps and other derivatives. The demand side consists of asset managers, leveraged hedge funds and other reportable. Asset managers are institutional investors and include pension funds, endowments, insurance companies, mutual funds and portfolio managers whose clients are predominantly institutional. Leveraged hedge funds are typically hedge funds and other types of money managers, including registered commodity trading advisors (CTAs); registered commodity pool operators (CPOs) or unregistered funds identified by CFTC. Their strategies may involve taking outright positions or arbitrage within and across markets. The traders may be engaged in managing and conducting proprietary futures trading and trading on behalf of speculative clients. Reportable traders that are not placed into one of the first three categories are placed into the other reportables category. The traders in this category mostly are using markets to hedge business risk, whether that risk is related to foreign exchange, equities or interest rates. This category includes corporate treasuries, central banks, smaller banks, mortgage originators, credit unions and any other reportable traders not assigned to the other three categories.

As seen in Figure A.4, asset managers make up the largest part of the market (when measured by open interest, i.e., the number of contracts outstanding) and are typically long in the market. Leveraged hedge funds are the second largest trader type on the demand side and they are typically short in the market. This can be due to speculation, but often it is for hedging purposes. The dealers are usually short in the market since they take the opposite side of trades. As seen in Figure A.6, the total number of registered dealer have remained constant throughout the sample period.

A. *LP Count: LOB versus CFTC*

Using CFTC data, we can compare the number of active LPs, N, to the number of registered CFTC LPs in the market. Figure A.6 plots the time series of N versus the CFTC LP number. In some periods, N is higher than the number of registered LPs, implying that other trader types must be posting limit orders. At other times, particularly during and following market slumps, N is lower than the number of CFTC LPs, meaning that some CFTC LPs choose not to provide liquidity. This has consistently been the case since 2018, and it confirms that LPs actively decide when to provide liquidity and when to step out of the market.

The intraday pattern of market depth closely resembles the pattern of trading volume, with liquidity lowest during Asian trading hours and highest during U.S. trading hours. The drop in depth and LP numbers has been almost uniform across the trading day, falling from an average of 133 LPs during U.S. trading hours in 2010–2014 to 24 in 2020–2024. Figure A.7 displays the full intraday patterns.

B. *Open Interest*

A potential explanation for why depth and the LP number have diminished recently, even when volatility is low, is that the demand for liquidity is simply lower. We can estimate this through the open interest, which counts the number of active futures contracts in the market. Figure A.3 shows that total open interest is around 3 million. It was slowly increasing from 2009 to 2018 and has been slightly decreasing since. However, this decrease is not nearly as large as the drop in market depth, and the open interest does not exhibit large downward jumps like the depth. Therefore, the open-interest-to-depth ratio is significantly lower today compared to before 2018.[19] It is not lower open interest that is associated with the decrease in market depth; on the contrary, it appears that open interest has decreased in response to the lower depth. Splitting the open interest by trader type, Figure A.4 shows that asset managers and leveraged funds have fewer outstanding contracts today. The drop in asset managers started in 2018, while leveraged funds have been decreasing since 2015. Surprisingly, the number of traders has not decreased as much and has even increased for asset managers.

A.5. Traditional Measures of Liquidity

A. *Lee and Ready (1991) algorithm*

Many of the liquidity measures described below require signed transactions. The literature standard is to follow the Lee and Ready (1991) algorithm. In the Lee and Ready (1991) algorithm, a trade is classified as a buy if the trade price is above the midpoint of the best bid/offer (BBO) and as a sell if the trade price is below the midpoint of the BBO.

B. *Adverse selection measures*

- Kyle (1985) lambda: price impact; measured as the coefficient λ_{it} in

$$\text{ret}_{itn} = \delta_{it} + \lambda_{it} S_{itn} + \epsilon_{itn},$$

[19] Specifically, the ratio of dollar depth at the best bid and ask to the notional value of open interest fell from approximately 0.6% in 2012 to 0.1% in 2022, a five-fold decline that far exceeds the modest decline in open interest itself.

where, for the nth five-minute period on date t and security i, ret_{itn} is the security return and S_{itn} is the signed square-root dollar volume:

$$S_{itn} = \sum_{k \in n} sign\left(\$vol_{itnk}\right) \sqrt{|\$vol_{itnk}|}$$

- Five minute price impact: gross losses to liquidity demanders due to adverse selection (Glosten and Harris, 1988); calculated as:

$$\text{Price impact}_{it} = \frac{\sum_{k \in t} \$vol_{itk} q_{itk} \left(\log\left(M_{it,k+5}\right) - \log\left(M_{itk}\right)\right)}{\sum_{k \in t} \$vol_{itk}},$$

where $q_{itk} = 1(-1)$ if is the kth trade is a buy (sell), M_{itk} is the midpoint of the prevailing BBO at the time of the kth trade, and $M_{it,k+5}$ is the midpoint of the prevailing BBO five minutes after the kth trade

- Trade-related variance of changes in efficient price/cumulative impulse response: price moves associated with recent trades are private information-based; as in Hasbrouck (1991a,b) estimate a VAR for returns and order flow:

$$\begin{aligned} r_{it\tau} &= \sum_{k=1}^{10} \alpha_{itk} r_{it,\tau-k} + \sum_{k=1}^{10} \beta_{itk} q_{it,\tau-k} + \epsilon_{r,it,\tau} \\ q_{it\tau} &= \sum_{k=1}^{10} \gamma_{itk} r_{it,\tau-k} + \sum_{k=1}^{10} \phi_{itk} q_{it,\tau-k} + \epsilon_{q,it,\tau}, \end{aligned}$$

where $r_{it,\tau}$ is the log return based on the BBO midpoint for security i on date t from trade time $\tau - 1$ to trade time τ. Invert to get the VMA representation:

$$\begin{bmatrix} r_{it\tau} \\ q_{it\tau} \end{bmatrix} = \begin{bmatrix} a_{it}(L) & b_{it}(L) \\ c_{it}(L) & d_{it}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{r,it,\tau} \\ \epsilon_{q,it,\tau} \end{bmatrix}$$

Then the variance of returns can be expressed as:

$$\sigma^2_{it,\omega} = \underbrace{\left(\sum_{j=0}^{\infty} b_{it,j}\right)^2 \sigma^2_{it,q}}_{\text{Trade-related}} + \underbrace{\left(\sum_{j=0}^{\infty} a_{it,j}\right)^2 \sigma^2_{it,r}}_{\text{Trade-orthogonal}}.$$

With this notation, we have:

  - Trade-related component of variance of changes in the efficient price (Hasbrouck, 1991b):

$$R_{it,\omega} = \frac{\left(\sum_{j=0}^{\infty} b_{it,j}\right)^2 \sigma^2_{it,q}}{\sigma^2_{it,\omega}}$$

– Cumulative impulse response (Hasbrouck, 1991a):

$$IR_{it,\omega} = \sum_{j=0}^{\infty} b_{it,j}.$$

- Amihud illiquidity:

$$\text{Illiquidity}_{it} = 1000 \frac{|r_{it}|}{\$vol_{it}},$$

where $\$vol_{it}$ is the unsigned volume on day t.

- Easley, Kiefer, O'Hara, and Paperman (1996) probability of informed trade (PIN). Informed traders receive private signals with probability α. If signal is positive (probability δ), buy orders from informed and uninformed arrive with Poisson intensity $\mu + \epsilon_B$; sell orders arrive from uninformed traders only with intensity ϵ_S. If signal is negative, reverse is true and buy orders arrive with intensity ϵ_B, sell orders arrive with intensity $\mu + \epsilon_S$. The parameters α, μ, ϵ_B, ϵ_S are estimated via MLE, with the likelihood given by

$$L\left(D_{PIN,it}\right) = L_{NI}\left(D_{PIN,it}\right) + L_{I^+}\left(D_{PIN,it}\right) + L_{I^-}\left(D_{PIN,it}\right),$$

where $L_{NI}\left(D_{PIN,it}\right)$ is the likelihood of observing B_{it} buys and S_{it} sells in a period without private information:

$$L_{NI}\left(D_{PIN,it}\right) = \left(1-\alpha_i\right) e^{-\epsilon_{B_i}} \frac{\epsilon_{B_i}^{B_{it}}}{B_{it}!} e^{-\epsilon_{S_i}} \frac{\epsilon_{S_i}^{S_{it}}}{S_{it}!},$$

$L_{I^+}\left(D_{PIN,it}\right)$ is the likelihood of observing B_{it} buys and S_{it} sells in a period with a positive private signal:

$$L_{I^+}\left(D_{PIN,it}\right) = \alpha_i \delta_i e^{-\left(\mu_i+\epsilon_{B_i}\right)} \frac{\left(\mu_i+\epsilon_{B_i}\right)^{B_{it}}}{B_{it}!} e^{-\epsilon_{S_i}} \frac{\epsilon_{S_i}^{S_{it}}}{S_{it}!},$$

and $L_{I^-}\left(D_{PIN,it}\right)$ is the likelihood of observing B_{it} buys and S_{it} sells in a period with a negative private signal:

$$L_{I^-}\left(D_{PIN,it}\right) = \alpha_i \left(1-\delta_i\right) e^{-\epsilon_{B_i}} \frac{\epsilon_{B_i}^{B_{it}}}{B_{it}!} e^{-\left(\mu_i+\epsilon_{S_i}\right)} \frac{\left(\mu_i+\epsilon_{S_i}\right)^{S_{it}}}{S_{it}!}.$$

The PIN of a security is then defined as the (conditional) probability of private information:

$$PIN_{it} \equiv \mathbb{P}\left(I_{it}=1 | D_{PIN,it}\right) = \frac{L_{I^+}\left(D_{PIN,it}\right) + L_{I^-}\left(D_{PIN,it}\right)}{L\left(D_{PIN,it}\right)}.$$

- Odders-White and Ready (2008) PIN (OWR): extends Kyle (1985) to allow for days with no private information. The OWR model uses three variables: the day's full order imbalance y_{et}, the intraday return r_{dt}, and the overnight return r_{ot}. The model is identified through the assumption that the covariance matrix of these three variables differs between days with private information arrival and days with only public information. Let α_i be the probability

that there is an information event on a given day, σ_{zi}^2 be the variance of the noise of the observed net order flow (y_{et}), σ_{ui}^2 be the variance of the net order flow from noise traders, σ_i^2 be the variance of the private signal received by the informed trader, $\sigma_{pd_i}^2$ be the variance of the intraday return, and $\sigma_{po_i}^2$ be the variance of the overnight return.

As with PIN, OWR is estimated via MLE, with the likelihood of observing $(y_{eit}, r_{dit}, r_{oit})$ on a day with and without an information event given by:

$$\begin{aligned} L_{I,it} &= \alpha_i \phi\left(y_{eit}, r_{dit}, r_{oit}; \Omega_{I,i}\right) \\ L_{NI,it} &= \alpha_i \phi\left(y_{eit}, r_{dit}, r_{oit}; \Omega_{NI,i}\right). \end{aligned}$$

Here, $\phi\left(\cdot; \Omega\right)$ is the joint normal pdf with mean 0 and covariance matrix Ω:

$$\Omega_{NI,i} = \begin{bmatrix} \sigma_{ui}^2 + \sigma_{zi}^2 & \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & -\sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 \\ \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & \sigma_{pd_i}^2 + \alpha_i\sigma_i^2/4 & -\alpha_i\sigma_i^2/4 \\ -\sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & -\alpha_i\sigma_i^2/4 & \sigma_{po_i}^2 + \alpha_i\sigma_i^2/4 \end{bmatrix};$$

$$\Omega_{I,i} = \begin{bmatrix} \left(1 + \alpha_i^{-1}\right)\sigma_{ui}^2 + \sigma_{zi}^2 & \alpha_i^{-\frac{1}{2}}\sigma_i\sigma_{ui}/2 + \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & \alpha_i^{-\frac{1}{2}}\sigma_i\sigma_{ui}/2 - \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 \\ \alpha_i^{-\frac{1}{2}}\sigma_i\sigma_{ui}/2 + \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & \sigma_{pd_i}^2 + \left(1 + \alpha_i\right)\sigma_i^2/4 & \left(1 - \alpha_i\right)\sigma_i^2/4 \\ \alpha_i^{-\frac{1}{2}}\sigma_i\sigma_{ui}/2 - \sqrt{\alpha_i}\sigma_i\sigma_{ui}/2 & \left(1 - \alpha_i\right)\sigma_i^2/4 & \sigma_{po_i}^2 + \left(1 + \alpha_i\right)\sigma_i^2/4 \end{bmatrix}.$$

The probability of an information event occurring on date t is then:

$$OWN_{it} = \frac{L_{I,it}}{L_{I,it} + L_{NI,it}}.$$

C. Market power

- Realized spread: revenue to liquidity providers assuming that the liquidity provider can close their position at the midpoint prevailing 5 minutes after the trade:

$$rspread_{it} = 2\frac{\sum_{k\in t} \$vol_{itk} q_{itk} \left(\log\left(P_{itk}\right) - \log\left(M_{it,k+5}\right)\right)}{\sum_{k\in t} \$vol_{itk}}.$$

D. Other measures

- Effective spread (Hasbrouck, 2010):

$$espread_{it} = 2\frac{\sum_{k\in t} \$vol_{itk} \left|\log\left(P_{itk}\right) - \log\left(M_{it,k+5}\right)\right|}{\sum_{k\in t} \$vol_{itk}}.$$

- Bid-ask spread:

$$baspread_{it} = 2\frac{ask_{it} - bid_{it}}{ask_{it} + bid_{it}},$$

where bid_{it} and ask_{it} are the prevailing bid and ask at the end of the period (e.g. at closing for intraday)

- Inverse limit order book slope (ILOBS) (Deuskar and Johnson, 2011): expected price impact; calculated as

$$ILOBS_{it\tau} = \frac{\sum_{k=1}^{K} \left(Mdist^2_{Ask,k,it\tau} + Mdist^2_{Bid,k,it\tau} \right)}{\sum_{k=1}^{K} \left(Mdist_{Ask,k,it\tau} \$vol_{Ask,k,it\tau} - Mdist_{Bid,k,it\tau} \$vol_{Bid,k,it\tau} \right)},$$

where $Mdist_{Ask,k,it\tau}$ is the absolute distance of the i limit order price on the ask side from the midprice and $Mdist_{Bid,k,it\tau}$ is the absolute distance of the i limit order price on the bid side from the midprice.

E. Traditional liquidity measures in the E-mini market

We now compare the LP number and market depth to other classical measures of (il)liquidity. The most common illiquidity measure is the bid–ask spread as it is directly measurable from prices and available in real time. Glosten and Milgrom (1985) show that the spread is increasing in asymmetric information, and liquidity providers therefore require a higher spread when market participants have private information. Unfortunately, it most often futile to use the spread for measuring liquidity of high volume exchange-traded contracts like the E-mini, because the minimum price change unit (the tick) acts as an effective minimum to the spread. The E-mini tick size is 0.25 index points = USD 12.5, and Figure A.5 shows that the spread is almost always equal to the tick, rendering it useless as a continuous liquidity measure.[20] Instead, the spread can be used as an indicator of large liquidity shocks, where a spread larger than 1 tick indicates a substantial illiquidity. Examples of this is seen in panel (b) of Figure A.5 where the spread increased to two ticks during the Global Financial Crisis in 2009. Additionally, we have more often experienced a higher spread following the COVID-19-related stock market crash in 2020 which is where depth and the LP number historically have been at their lowest.

Besides the spread, we also consider Kyle's lambda (see Kyle (1985)) and Amihud's illiquidity measure (see Amihud (2002)) which both measure market impact, i.e., how trading activity affects price movements. Panels (d) and (f) of Figure A.5 show that market impact has indeed been higher since 2020 which is when depth and the LP number is particularly low. A drawback of Amihud's illiquidity measure and Kyle's lambda compared to market depth and the LP number is that they are not measurable in real time; they require a period to be measured over. This is disadvantageous to traders in high frequency markets, where conditions can change almost instantaneously. Furthermore, price returns are used to computed the price impact measures, making them non-eligible to be used for explaining returns.

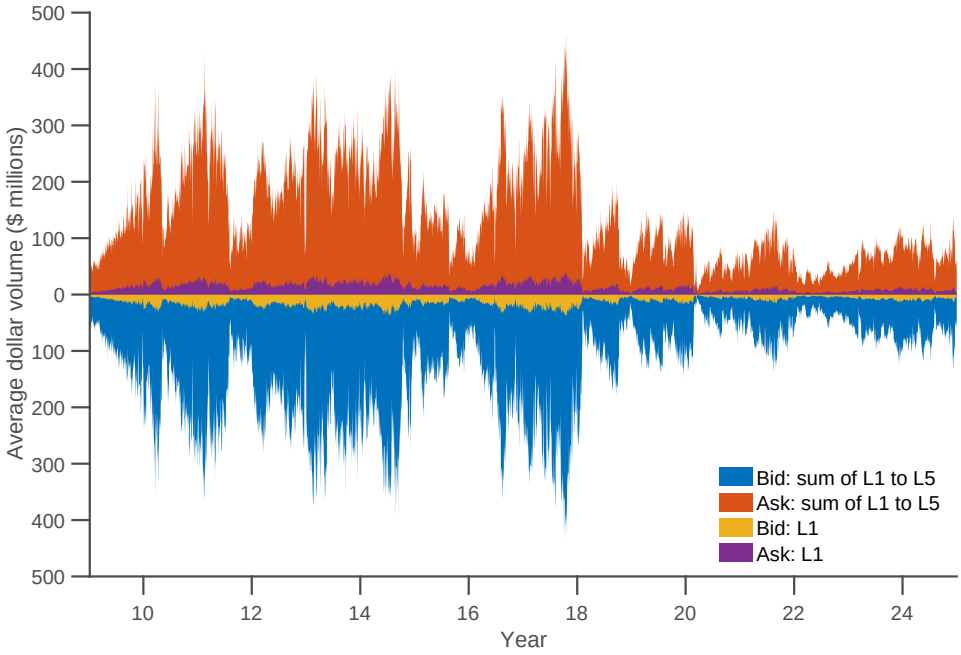
F. LP number versus traditional liquidity measures

Market depth is by definition almost equal to the standard definition of liquidity: when the LOB is deep, market participants can instantaneously buy or sell a large amount of the asset without affecting its price. As discussed, the long-run depth of the E-mini is driven by the LP number, suggesting that the LP number is a fundamental component of market liquidity. This is confirmed in Panel (a) of Table A.1 which shows that depth explains 76% of variation in the LP number (and vice versa). The table also shows that the traditional liquidity measures have low explanatory power on the LP number, suggesting that the LP number contains information on market fundamentals

[20] We note that while the tick-level spread is uninformative as a continuous measure, the *relative* spread–measured in basis points–declined mechanically over the sample period. As the S&P 500 index rose from approximately 900 in 2009 to over 6,000 by 2024, the one-tick spread fell from roughly 1.4 basis points to approximately 0.2 basis points in relative terms.

not captured by conventional liquidity measures. To quantify this, Panel (b) in Table A.1 considers volatility regressed on the various liquidity measures. All measures have explanatory power in simple regressions, and regression (vi) shows that the LP number adds incremental explanatory power for volatility beyond that provided by the common liquidity measures, consistent with the LP number capturing a distinct dimension of market liquidity. Specifically, the R^2 rises from 47.8% in regression (v), which includes Kyle's λ, the Amihud measure, and the bid-ask spread, to 51.8% in regression (vi) when N is added–an increase of 4.0 percentage points. The partial R^2 attributable to the LP number, computed as the fraction of residual variance explained, is 7.6%, indicating that LP participation conveys economically meaningful information about volatility beyond what traditional liquidity proxies capture.

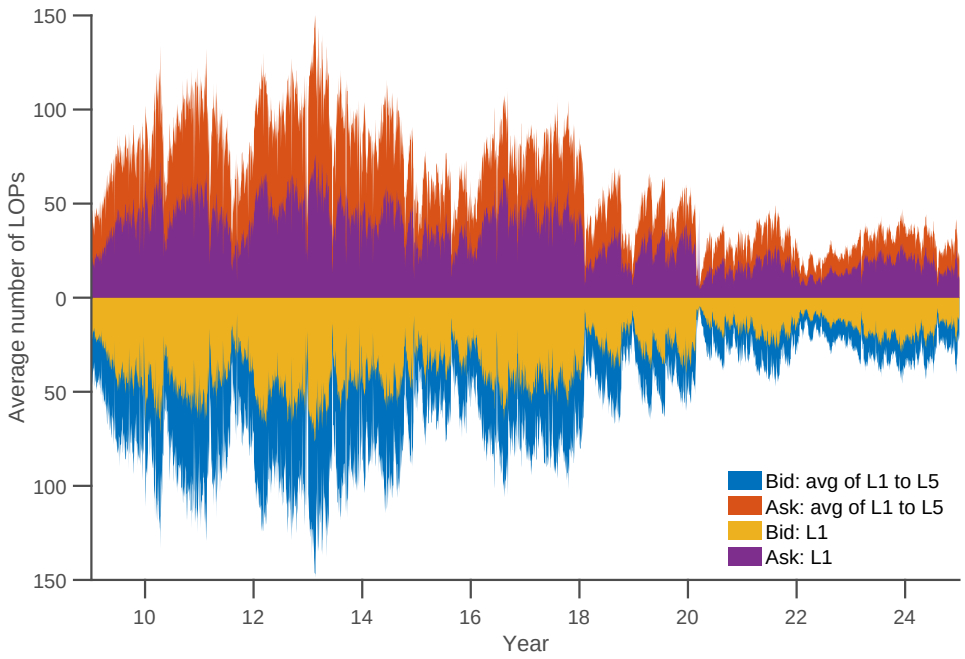
A.6. Figures



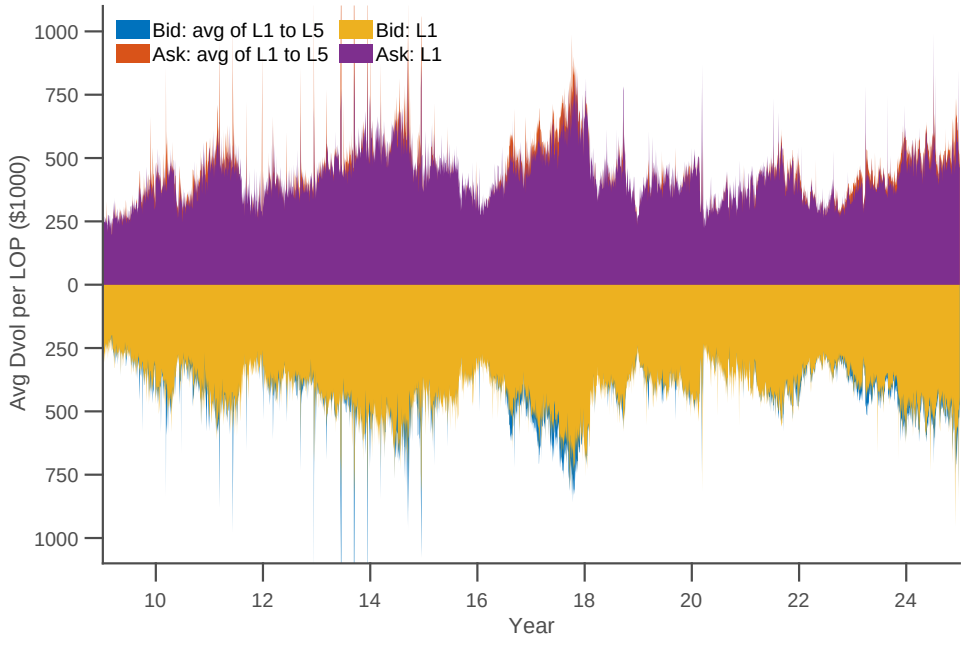
(a) both sides or the LOB

Figure A.1. Time series of market depth

The figure plots the time series of market depth measured as dollar volume available at level 1 and at level 1 to 5 of the limit order book for the S&P 500 E-mini futures contract. The buy and sell side are identical. The daily dollar volume is measured as averages of 5-minute samples throughout the trading day. The sample period is January 2009 to December 2024.



(a) Number of LOPs



(b) Depth per LOP

Figure A.2. Time series of LOP number and depth per LOP

Figure (a) plots the time series for the number of LOPs at level 1 of the order book and for the average number of LOPs at level 1 to 5. Figure (b) plots the time series of the average daily depth offered per LOP. Depth is measured as the average dollar volume offered per LOP. The daily LOP number and dollar volume per LOP are measured as averages of 5-minute samples throughout the trading day. The sample period is January 2009 to December 2024.

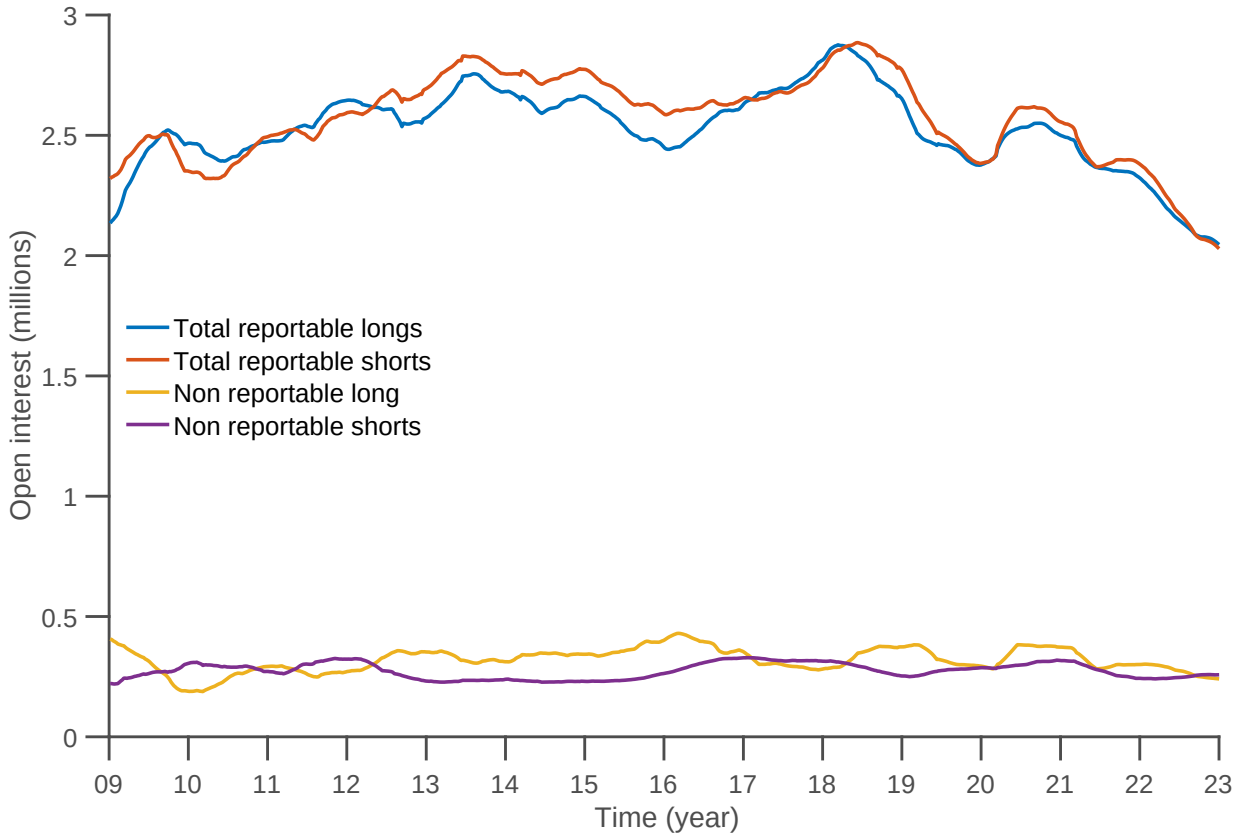
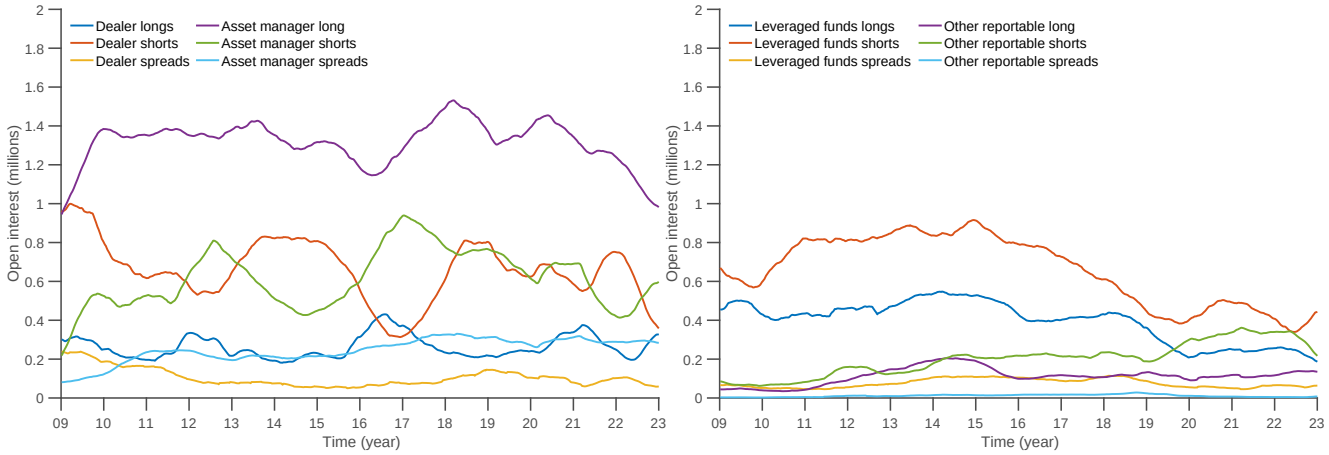


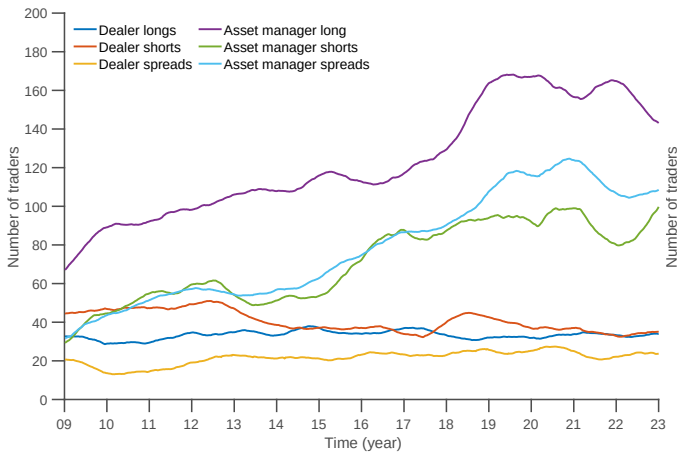
Figure A.3. Time series of open interest

The figure plots the time series of the open interest for the S&P 500 E-mini futures contract. Open interest is the total number of futures contracts held by market participants at the end of the trading day. Open interest is calculated by adding all the contracts from opened trades and subtracting the contracts when a trade is closed. Open interest and trading volume are related concepts, with the key difference being that volume counts all contracts that have been traded, while open interest is a total of contracts that remain open in the market. Traders interpret open interest as the cash flowing to the market: as open interest increases, more money is moving into the futures contract, and as open interest declines money is moving out of the market. The sample period is January 2009 to December 2022.

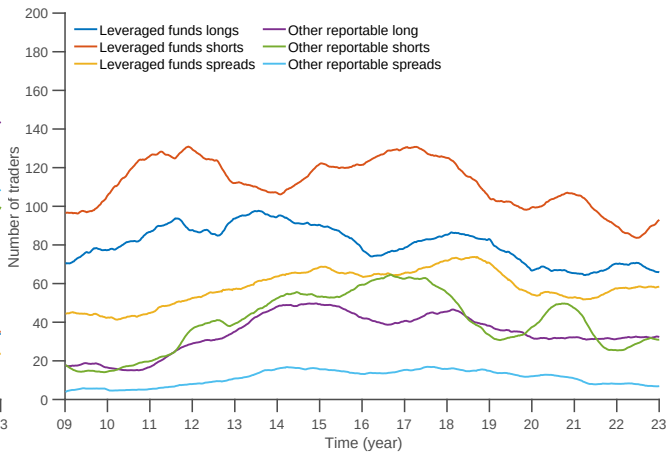


(a) Open interest by trader type

(b) Open interest by trader type

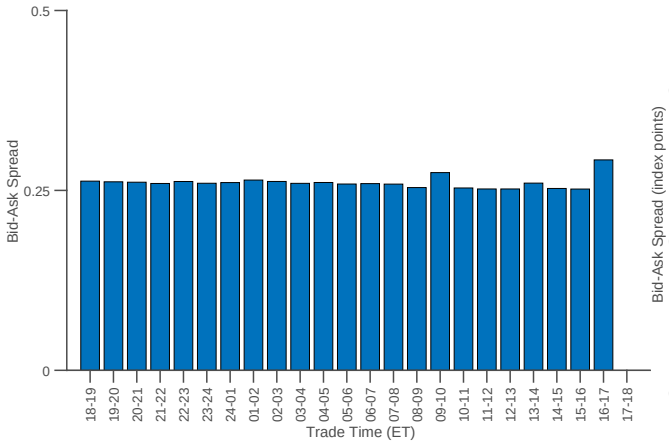


(c) Number of traders by trader type

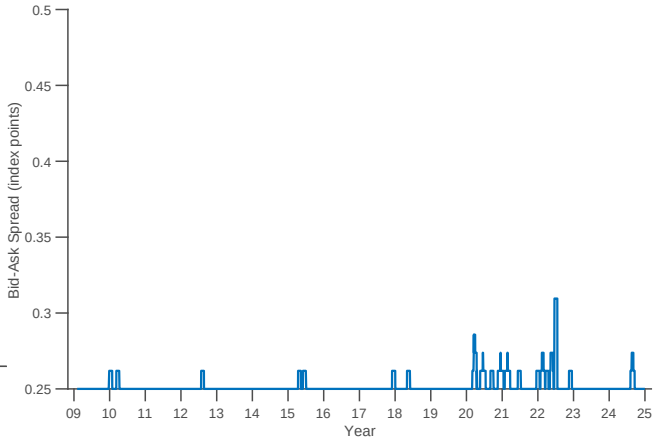


(d) Number of traders by trader type

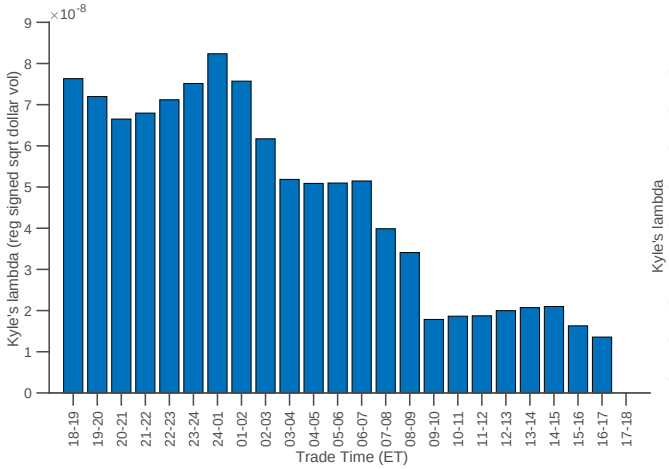
Figure A.4. Time series of open interest by trader type and number of traders
The figure plots the time series of the open interest and number of traders by type for the S&P 500 E-mini futures contract. The sample period is January 2009 to December 2024.



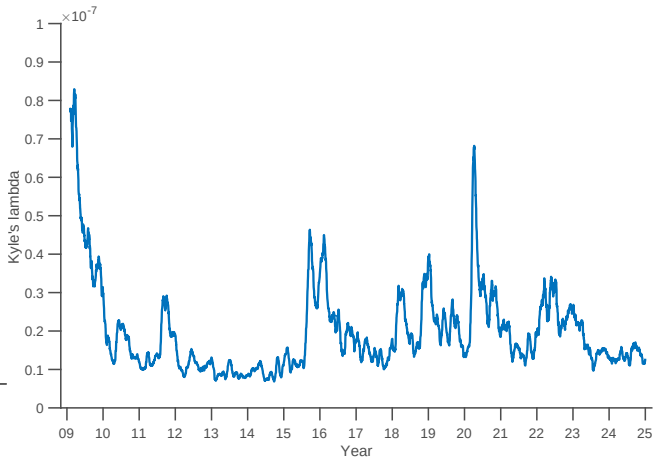
(a) bid–ask spread - intraday hourly average



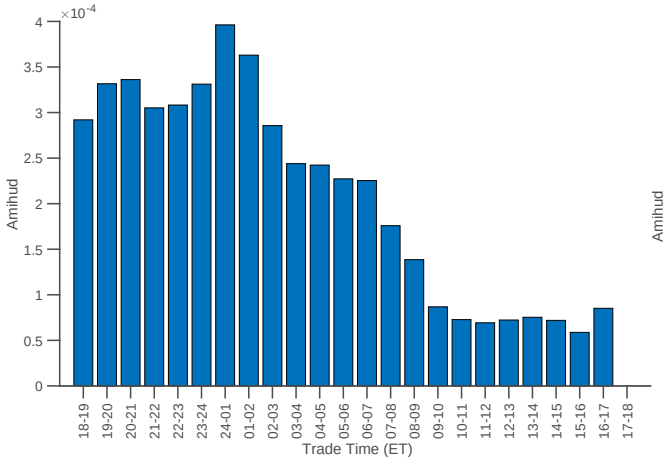
(b) bid–ask spread - time series



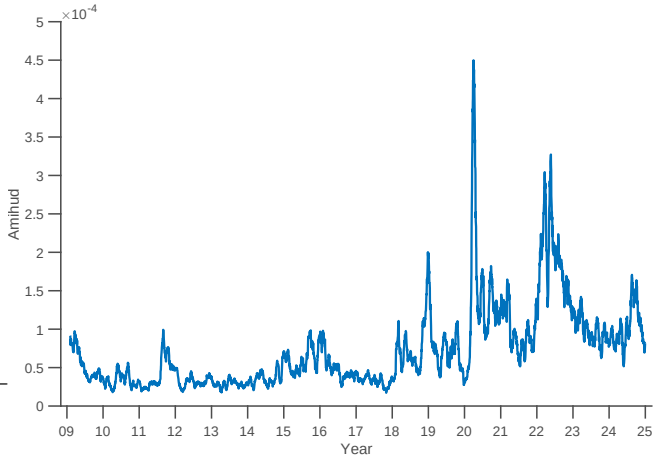
(c) Kyle's lambda - intraday hourly average



(d) Kyle's lambda - time series



(e) Amihud - intraday hourly average



(f) Amihud - time series

Figure A.5. Traditional liquidity measures

Figure (a) and (b) plots the bid–ask spread. (c) and (d) plots Kyle's lambda, and (e) and (f) plots Amihud's illiquidity measure defined as $1/N \sum_{i=1}^{N} |r_i|/\$V_i$, where N is the number of observations, r_i is the return and $\$V$ is the dollar trading volume. Kyle's lambda and Amihud are computed at an hour frequency based of 1-minute sampling. Plots of intraday averages show the averages of a given trading hour across all trading days in the sample period. All time series are plotted as rolling monthly averages of daily estimates to reduce noise. The sample period is January 2009 to December 2024.

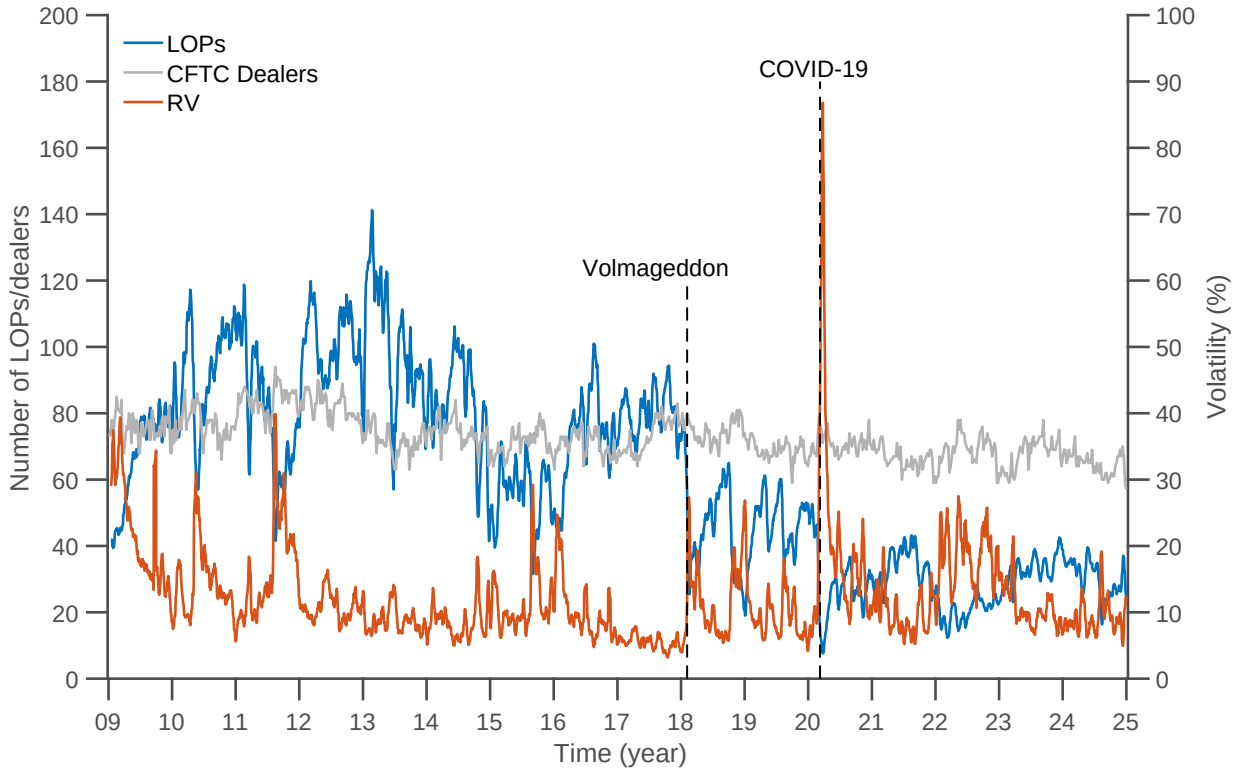
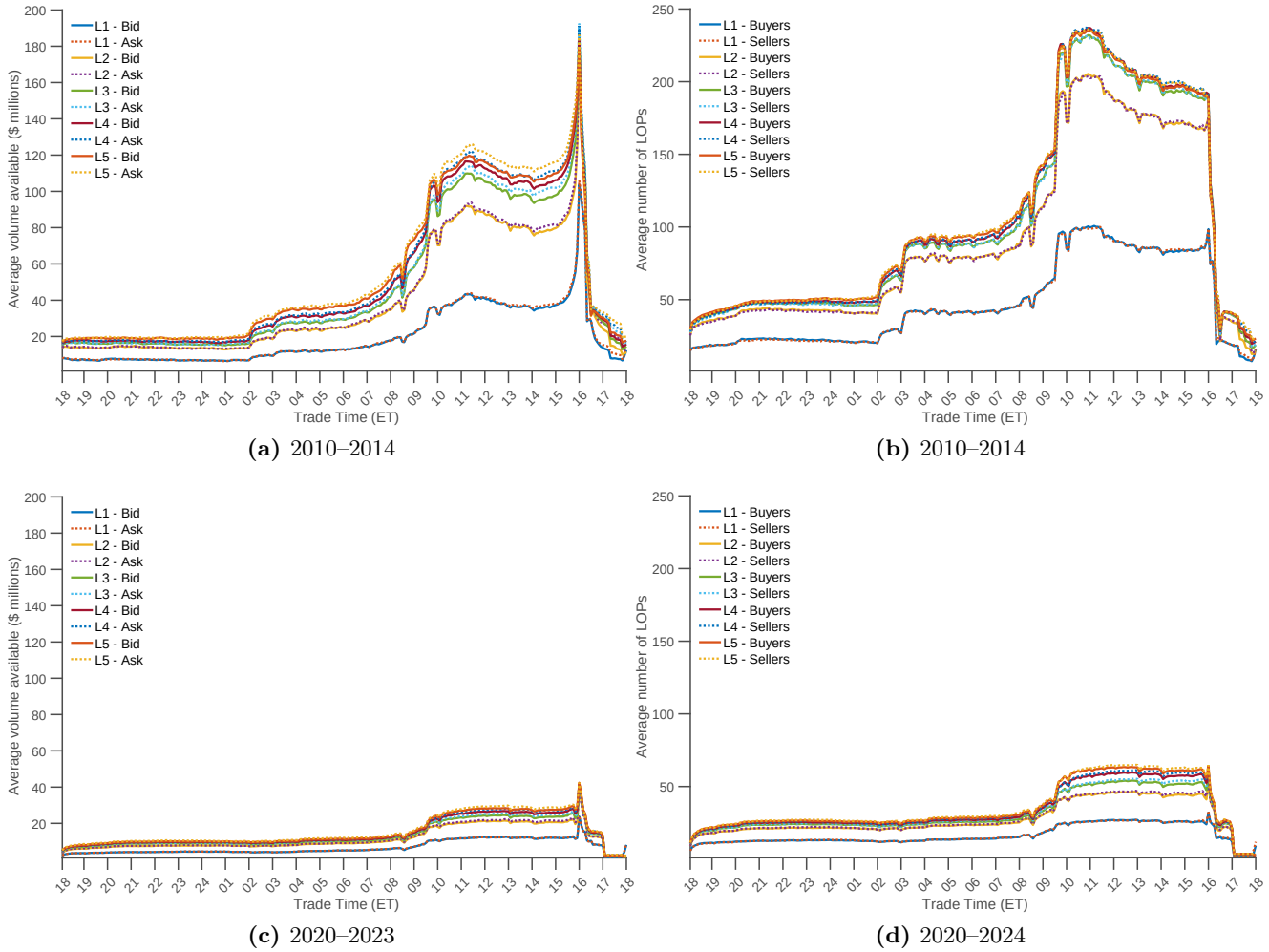


Figure A.6. Time series of LOPs, CFTC dealers, and volatility

The figure plots the time series of 1) the average daily number of LOPs across level 1 to 5 of the limit order book for the S&P 500 E-mini futures contract, 2) the number of registered CFTC dealers (computed as long plus short positions), and 3) the (annualized) daily realized volatility computed from 5-minute log returns. All variables are computed as a 10 day rolling average to reduce noise. The sample period is January 2009 to December 2024.



(a) 2010–2014 **(b)** 2010–2014

(c) 2020–2023 **(d)** 2020–2024

Figure A.7. Intraday depth and LOP number

Figure (a) and (c) plots the intraday average market depth measured as dollar volume available for level 1 to 5 of the order book for the S&P 500 E-mini futures contract. Figure (b) and (d) plots the intraday average number of LOPs for level 1 to 5 of the order book. The dollar volume and LOP number are measured at a 5 minute frequency throughout the trading day. The sample period is January 2010 to December 2014 for (a) and (b) and April 2020 to December 2024 for (c) and (d).

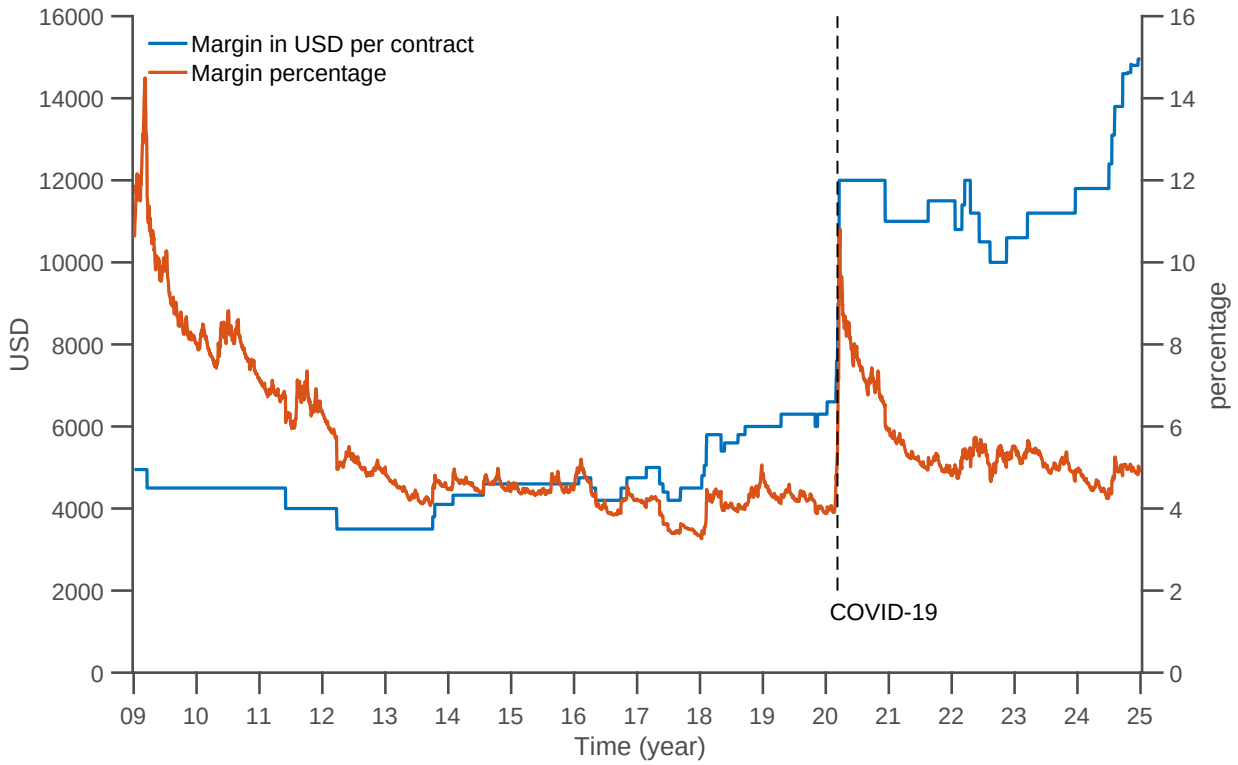


Figure A.8. Time series of futures margin

The figure plots the time series of the margin for the S&P 500 E-mini futures contract. The margin is shown both in U.S. dollars per contracts, and in percentage of the contract value, i.e., the margin relative to the current value of the E-mini contract. The sample period is January 2009 to December 2022.

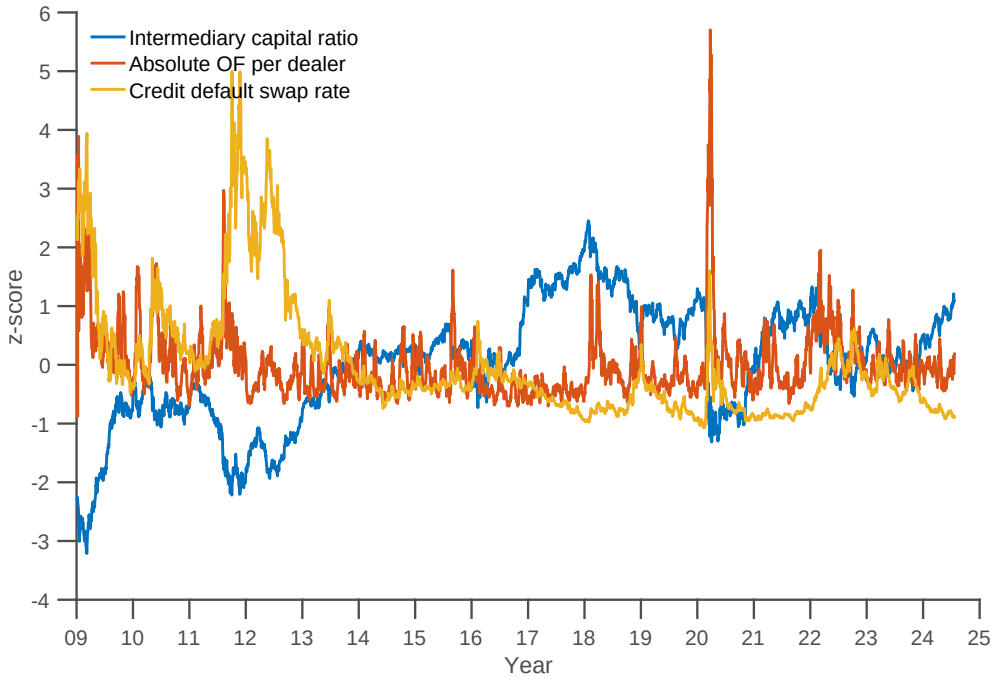


Figure A.9. Time series of LOP constraint proxies

The figure plots time series of our proxies for the LOP constraints: The capital ratio of intermediaries, the daily absolute order flow per LOP, and the average credit default swap rate of G14 dealers. All variables are z-scored. The sample period is January 2009 to December 2024.

A.7. Tables

	(i)	(ii)	(iii)	(iv)	(v)
depth	0.89				0.88
	(0.09)				(0.082)
Kyle's λ		-0.10			0.23
		(0.21)			(0.064)
Amihud illiq			-0.17		-0.041
			(0.12)		(0.026)
BA spread				-0.33	-0.19
				(0.17)	(0.028)
$R^2(\%)$	78.39	1.01	2.90	10.56	81.90

(a) N_t as response

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
N_t	-0.34					-0.21
	(0.09)					(0.072)
Kyle's λ		0.64			0.46	0.50
		(0.33)			(0.27)	(0.25)
Amihud illiq			0.40		-0.028	-0.040
			(0.12)		(0.066)	(0.061)
BA spread				0.59	0.33	0.24
				(0.095)	(0.13)	(0.14)
$R^2(\%)$	11.26	41.18	16.11	34.51	47.81	51.77

(b) σ_t as response

Table A.1. Traditional liquidity measures versus the LOP number and volatility
This table considers the daily number of limit order providers (LOPs), N_t, and volatility, σ_t, regressed on traditional liquidity measures, LIQ_t^i, for the S&P 500 E-mini futures contract. The regressions are in levels.

$$N_t = \alpha + \beta LIQ_t^i + \epsilon_t \qquad \text{and} \qquad \sigma_t = \alpha + \beta LIQ_t^i + \epsilon_t.$$

Liquidity is measured using (i) the market depth (measured as the available dollar volume), (ii) Kyle's λ, (iii) Amihud's (2002) illiquidity measure, and (iv) the bid-ask spread. All variables are z-scored. Standard errors are estimated using a block bootstrap with 10,000 draws and where the optimal block size is chosen following Patton (2009). The sample period is January 2009 to December 2024.

Hour	09-10	10-11	11-12	12-13	13-14	14-15	15-16	CTC	CTO	OTC
intercept	0.27	0.63	0.96	0.34	−0.37	0.87	−0.56	4.19	−0.47	3.46
se	0.44	0.45	0.39	0.31	0.32	0.38	0.48	1.48	0.87	1.10
SV	2.23	2.55	2.06	1.76	1.84	2.08	2.32	5.54	4.35	4.71
se	0.05	0.06	0.07	0.09	0.04	0.06	0.10	0.30	0.13	0.15
R^2	30.35	35.68	32.16	29.70	31.80	31.66	25.75	22.39	27.93	27.67

(a)

Hour	09-10	10-11	11-12	12-13	13-14	14-15	15-16	CTC	CTO	OTC
intercept	−7.56	−6.28	5.88	−3.37	−10.41	0.95	−0.88	−24.81	−20.05	−18.51
se	4.73	4.78	3.68	3.24	4.13	3.80	5.43	13.60	9.68	10.43
SV	3.78	5.12	3.85	5.19	4.77	6.70	8.17	7.15	14.79	6.17
se	0.62	0.61	0.65	1.14	0.55	0.61	1.04	2.03	1.45	1.45
N	1.77	1.49	−1.05	0.82	2.21	0.01	0.13	7.20	5.31	4.81
se	1.02	0.98	0.75	0.66	0.86	0.79	1.13	3.16	2.53	2.14
$SV \times N$	−0.35	−0.55	−0.38	−0.74	−0.63	−1.00	−1.27	−0.40	−2.85	−0.32
se	0.13	0.12	0.13	0.23	0.11	0.13	0.22	0.47	0.38	0.30
R^2	30.71	36.44	32.62	31.63	33.51	34.39	28.49	22.51	30.82	27.80

(b)

Table A.2. Returns, order imbalance and the number of liquidity providers

This table shows coefficients for regressions of the types $R_{t,h} = \alpha + \beta_1 OF_{t,h} + \epsilon_t$ and $R_{t,h} = \alpha + \beta_1 OF_{t,h} + \beta_2 N_{t,h} + \beta_3 OF_{t,h} \times N_{t,h} + \epsilon_t$, where $R_{t,h}$ is the return measured in basis points over a given trading period, h, on day t, $OF_{t,h}$ is order flow in contracts/1000, and $N_{t,h}$ is the average number of market participants posting limit orders across the first five levels of the limit order book. Standard errors are estimated using a block bootstrap with 10,000 draws, where the optimal block size is chosen following Patton (2009). The sample period is January 2009 to December 2024.

	2	3	4	5	6	7	8
1: ΔN	88.	−10.	−17.	−33.	−70.	−34.	5.
2: Δ Depth		−22.	−18.	−29.	−57.	−42.	3.
3: Δ Kyle's λ			56.	59.	−13.	64.	0.
4: Δ Amihud illq				52.	9.	40.	−3.
5: Δ B/A spread					25.	59.	−3.
6: Δ trading vol						28.	−15.
7: $\Delta\sigma$							−13.
8: Return							

(a) Levels

	2	3	4	5	6	7	8
1: ΔN	62.	−14.	−7.	−13.	−16.	−30.	21.
2: Δ Depth		−10.	−6.	−10.	−15.	−22.	11.
3: Δ Kyle's λ			64.	75.	−5.	67.	−1.
4: Δ Amihud illq				48.	−3.	46.	−5.
5: Δ B/A spread					−2.	57.	−1.
6: Δ trading vol						40.	−27.
7: $\Delta\sigma$							−21.
8: Return							

(b) Differences

Table A.3. Pairwise Correlations

This table shows the pairwise correlations (in %) of both the daily levels and the daily changes of liquidity measures, volatility and prices. The sample period is January 2009 to December 2024.