Instant Liquidity, Latent Risk *

Lionel Melin† Florent Rouxelin‡

August 29, 2025

Abstract

We develop a continuous-time model of segmented bond markets to examine the liquidity benefits and latent risks introduced by exchange-traded funds (ETFs). Our model shows that ETFs provide liquidity hedging, compress OTC bond yield spreads, and improve price discovery—especially in illiquid markets (as when OTC trading froze during the COVID-19 crisis). Yet our model also reveals spillovers, path-dependent equilibria, and the risk of redemption-driven markdowns, which fire sales could fuel into NAV spirals, leading to stress amplification. Our results inform ETF design and regulation to best harness liquidity while mitigating instability under stress.

JEL Codes: G12, G14, G18, G28

Keywords: exchange-traded funds (ETFs), fixed-income ETFs, collateralized loan obligations (CLOs), municipal bonds (munis), financial intermediation, systemic risk, market segmentation, liquidity risk

*We thank Fernando Alvarez, Veronica Guerrieri, John Huizinga, Riccardo Lagos, Guillaume Rocheteau, Robert Shimer, and Pierre-Olivier Weill for comments and feedback on the theoretical model in this paper.

†MacroLucid Research. Email lionel.melin@macrolucid.com

‡Assistant Professor of Finance, Department of Finance, College of Business, Florida International University. Email frouxelin@fiu.edu (corresponding author)

1 Introduction

Could the rapid rise of fixed-income exchange-traded funds (ETFs), particularly those holding illiquid or opaque credit exposures, be sowing the seeds of a future liquidity crisis akin to 2007–2008? These products offer daily tradability in assets that seldom trade and now span a vast footprint across credit markets, from investment-grade and high-yield corporate bonds to U.S. Treasuries, mortgage-backed securities (MBS), municipal bonds, emerging-market sovereign debt, floating-rate loans, and more recently, structured assets such as collateralized loan obligations (CLOs). Fixed-income ETFs worldwide managed roughly $2.6 trillion in assets as of early 2025[1], a dramatic increase over the past decade. Within this landscape, CLO ETFs are a relatively small segment of the market but offer a striking illustration of how ETF structures can wrap relatively inert, thinly traded instruments into vehicles that appear liquid[2]. Just as complex loans were once repackaged into benign-seeming securities before the Global Financial Crisis (GFC), today's ETFs on illiquid assets may be introducing systemic vulnerabilities through a new route: a liquidity mirage created by the ETF wrapper. The underlying mechanism, transforming non-fungible, low-turnover exposures into continuously traded shares, bears an uneasy resemblance to the pre-GFC securitization chains that collapsed when market liquidity vanished (Coval et al., 2009; Gorton and Metrick, 2012). The mismatch between the apparent tradability of ETF shares and the inertia of their underlying assets introduces a fragility that, if put under stress, could reverberate across markets.

Recent market episodes have offered glimpses of this fragility. The COVID-19 crisis in March 2020 is a case in point: as credit markets seized up, many corporate bond ETFs traded at deep discounts to their net asset values (NAVs), with some large funds falling more than 5% below NAV amid a panicked dash for cash. Although these dislocations eventually corrected, it required extraordinary intervention. In an unprecedented move, the U.S. Federal Reserve stepped in as a backstop in March 2020, establishing facilities to purchase corporate bond ETFs and restore market functioning (Kargar et al., 2021).[3] This intervention, described by some as the Fed acting as a "market maker of last resort," underscored how ETF liquidity mismatches can transmit and amplify stress. It also marked the first time a central bank directly supported ETFs, highlighting the vehicles' growing importance and potential to malfunction

[1]BlackRock. (April 2025). *Bond Market Opportunities Meet ETF Innovation*. Available at: ishares.com/etf-innovation

[2]S&P Global Ratings, "ABS Frontiers: How the Burgeoning CLO ETF Sector Could Impact the Broader CLO Market," November 26, 2024. Available at: spglobal.com/ratings/clo-etf-frontiers

[3]For a policy perspective emphasizing the novelty of this intervention and its implications for future central bank roles in market functioning, see Hauser (2021). *From lender of last resort to market maker of last resort: why central banks need new tools for dealing with market dysfunction*. Bank of England speech, 7 January. Available at: bankofengland.co.uk/speech/hauser-2021

in modern credit markets.

Originally launched as low-cost vehicles for equity exposure, ETFs now intermediate credit risk through a structure that couples a liquid, exchange-traded share market with an underlying over-the-counter (OTC) market for the portfolio assets. Authorized participants (APs) create ETF shares by delivering a specified basket of securities to the fund sponsor, and redeem shares by receiving a basket of underlying assets in return (Ben-David et al., 2018, see Appendix A for operational details). In fixed-income ETFs, these baskets are not pro-rata slices of the portfolio but custom-built, allowing asymmetries in what is delivered or received. The costs of transferring assets between venues can differ substantially between creations and redemptions, especially when some bonds are illiquid. This segmentation naturally splits investors: those seeking immediacy trade ETF shares, while others remain in OTC markets. ETFs thereby offer real-time tradability over inherently sluggish instruments, extending market access—but at the risk of masking underlying liquidity constraints.[4]

To formalize these dynamics, we construct a continuous-time model of dual-market trading, where investors allocate between an illiquid OTC asset and a liquid claim traded on a centralized exchange (CEX), representing ETF shares. Investor preferences evolve stochastically, prompting rebalancing subject to trading frictions. Intermediaries (analogous to APs) endogenously transform assets between platforms by creating or redeeming ETF shares, subject to direction-specific transfer costs. These mechanisms generate endogenous market segmentation, dynamic arbitrage activity, and feedback loops between ETF and OTC pricing.

Our framework fills a critical gap in the literature by providing the first continuous-time model that captures both dynamic market sizes and trading in an illiquid OTC market and a liquid CEX venue. Previous strands of research have highlighted important mechanisms in isolation: OTC search and bargaining (Duffie et al., 2005, 2007; Gârleanu, 2009; Lagos and Rocheteau, 2007), dealer-inventory models of liquidity provision (Amihud and Mendelson, 1980; Gârleanu and Pedersen, 2016; Lester et al., 2015), and seminal work on market freezes, adverse selection, and venue competition in segmented markets (Chiu and Koeppl, 2016; Liu, 2016; Pagnotta and Philippon, 2018). Related research has also examined ETF-specific phenomena such as arbitrage and mispricing (Brown et al., 2021), fire-sale propagation (O'Hara and Zhou, 2021), and price informativeness (Glosten et al., 2021), but none integrate these components in a unified dynamic framework. We extend classic OTC models by incorporating dual-market clearing and arbitrage flows, in which intermediaries not only manage inventories but also transfer liquidity across venues. Their incentives and market power (Glode

[4]In the ETF space, research has primarily focused on equity vehicles.Lettau and Madhavan (2018) assess structural fragility, Marta and Riva (2025) examine whether ETFs amplify volatility or comovement in underlying assets, and Glosten et al. (2021) study the ability of ETF prices to reflect fundamental information.

and Opp, 2020; Glode et al., 2022) play a key role in determining spreads, segmentation, and resilience.

Within this framework, our analysis makes three contributions. First, we develop a structural continuous-time model of segmented markets with endogenous ETF issuance and redemptions. Second, we simulate the model to replicate key empirical phenomena, including pricing dislocations, liquidity spirals, and arbitrage breakdowns. Third, we provide policy insights on ETF design features that modulate systemic risk under stress.

Our first contribution is theoretical. Building on the model described above, we derive closed-form equilibrium expressions for prices, spreads, and investor welfare, which allow us to study how liquidity provision and segmentation evolve endogenously. The framework extends the work of Lagos and Rocheteau (2009) by introducing platform-linked state variables that connect the OTC and CEX venues and by allowing ETF liquidity capacity to adjust dynamically through arbitrage.[5] These mechanisms make it possible to analyze how stress and shocks reshape segmentation and liquidity in real time. By embedding cross-venue frictions directly into equilibrium, the model highlights the structural sources of systemic vulnerability when liquidity transformation mechanisms weaken under stress.

Our second contribution is empirical. We calibrate and simulate the model to examine how liquidity and prices respond under varying frictions and trading conditions. The model rationalizes several observed patterns. First, ETF trading compresses OTC spreads and enhances market liquidity, especially in less accessible segments—consistent with evidence in Holden and Nam (2024). Second, when OTC trading halts, ETFs continue to reflect information and often lead bond prices, consistent with the "speed premium" in Üslü (2019). Third, the model captures persistent ETF discounts during stress periods, such as March 2020, and reproduces conditions under which arbitrage becomes ineffective (Holden and Nam, 2024; Pinter et al., 2024). These outcomes are also consistent with dislocations observed during the Taper Tantrum (Dannhauser and Hoseinzade, 2021) and with broader theoretical mechanisms of liquidity spirals and systemic fragility (Acharya et al., 2017; Brunnermeier and Pedersen, 2009).

The third contribution concerns policy. Our model highlights the feedback loop between redemption frictions and stress amplification. When redemption costs rise or authorized participants (APs) withdraw, ETF prices diverge from NAV and become driven by imbalances in secondary-market trading. These frictions interact with investor flows and valuation transparency to generate self-reinforcing dislocations, akin to run dynamics (Shleifer and Vishny, 1997). Modest changes to redemption protocols or arbitrage incentives can significantly improve resilience without undermining ETF efficiency. These findings support current calls for

[5] Melin (2012) first developed a framework with coexisting OTC and exchange.

enhanced transparency and stress-testing of ETF structure (Grill et al., 2018).

In summary, our study provides a unified theoretical and empirical account of the rise of fixed-income ETFs and the structural vulnerabilities that accompany their liquidity transformation function. The remainder of the paper is organized as follows. Section 2 presents empirical evidence on fixed-income ETFs, emphasizing their growth, segmentation, and pricing dislocations that motivate our model. Section 3 develops the continuous-time model in detail and characterizes equilibrium across the OTC and CEX platforms. Section 4 presents simulation results, illustrating how shocks propagate between markets, and explores comparative statics under varying frictions. Section 5 discusses the implications for liquidity mismatch, evaluates potential design and regulatory interventions, and draws lessons for systemic risk management. Section 6 concludes.

2 Empirical Motivation

The rapid growth of fixed-income ETFs has redefined access to credit markets by packaging traditionally illiquid exposures into continuously tradable shares. This evolution, shown in Figure 1, spans multiple asset classes including government, corporate, municipal, and structured credit. Particularly since 2020, securitized and tax-exempt bonds have increasingly migrated into the ETF wrapper, introducing liquidity mechanisms atop otherwise segmented markets.[6] As ETFs intermediate more opaque credit exposures, understanding the degree to which their prices align with NAV becomes central to assessing market integrity and liquidity transformation.

Government bond ETFs provide a benchmark for efficient arbitrage. These funds, which account for approximately half of the fixed-income ETF universe,[7] exhibit near-zero premium-discount behavior throughout the sample period. Figure 2 illustrates this alignment even during episodes of market stress, such as March 2020 and the subsequent policy normalization. These ETFs typically hold U.S. Treasury securities with high transparency, deep liquidity, and reliable pricing. The arbitrage mechanism remains active in this segment, reinforced by primary-market clarity and high secondary-market turnover. For context, the figure also overlays the 90-day Secured Overnight Financing Rate (SOFR), a benchmark proxy for institutional funding conditions.[8]

[6] The detailed construction of the dataset, filtering rules, and classification procedures are described in Appendix B. The data are sourced from the Center for Research in Security Prices (CRSP).

[7] Government bond ETFs are classified using the CRSP Mutual Funds database. Full construction details are provided in Appendix B.

[8] SOFR data are sourced from the Federal Reserve Economic Data (FRED) database, maintained by the Federal Reserve Bank of St. Louis (FRED series: SOFR90DAYAVG).

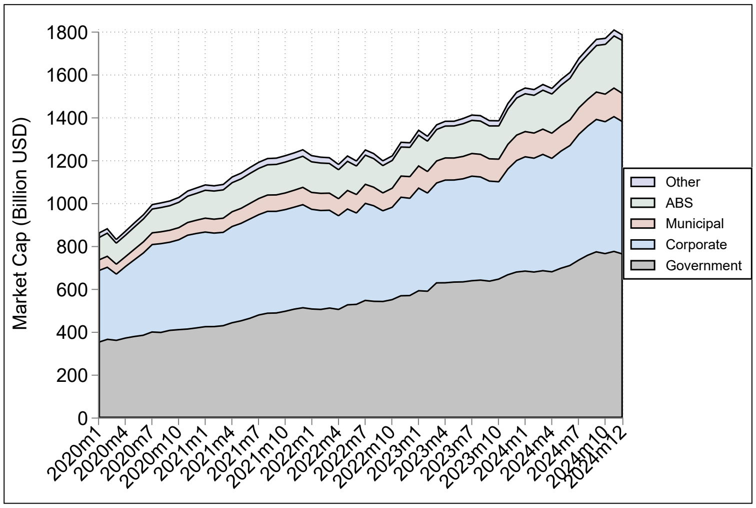


Figure 1: Fixed-Income ETF Market Capitalization by Asset Class, 2020–2024. This figure plots the monthly market capitalization for five categories of fixed-income ETFs: municipal, corporate, government, ABS, and other fixed-income ETFs. The data are aggregated at the asset-class level using the CRSP Mutual Funds database. Market capitalization is reported in billions of USD and spans the period from January 2020 through December 2024.

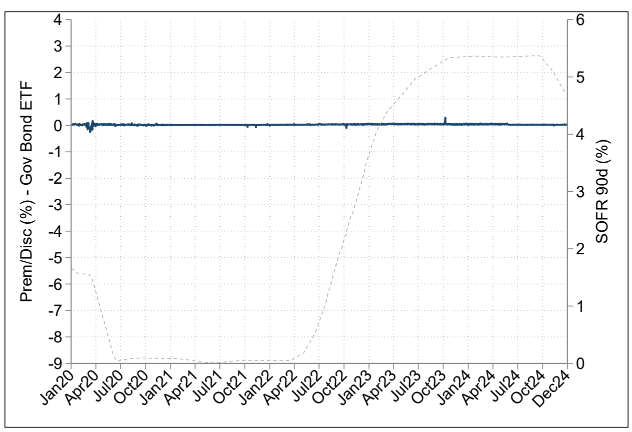


Figure 2: Government Bond ETF Premium-Discount and 90-Day SOFR, 2020–2024. This figure plots the daily premium or discount (in percent) of a representative set of ETFs providing exposure to U.S. government bonds, alongside the 90-day SOFR, from January 2020 through December 2024.

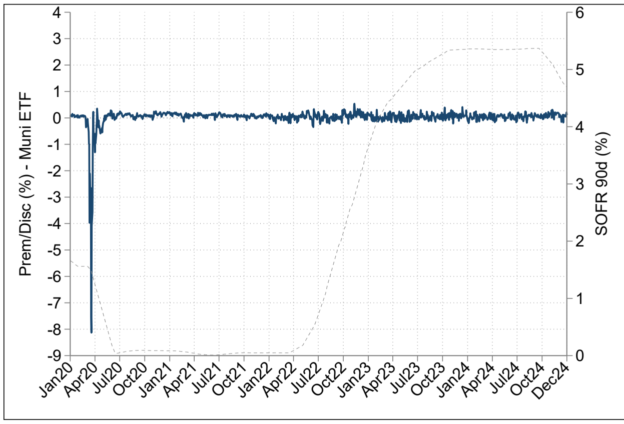
Outside of this benchmark segment, however, ETF pricing exhibits recurring dislocations. Municipal bond ETFs display the most persistent deviations from NAV. Figure 3 shows discounts approaching 8 percent in March 2020 and repeated disruptions even in calm periods. These deviations reflect a form of segmentation in the municipal bond market that operates alongside, rather than within, the OTC/CEX divide: tax-exempt cash flows, thin trading, and limited institutional participation constrain arbitrage activity. As documented by Schwert (2017), frictions in pricing and dealer intermediation reduce the responsiveness of muni bonds to new information. Moreover, market structure data show that households, either directly or through mutual funds and ETFs, held approximately 66 percent of outstanding municipal securities as of early 2022 (Municipal Securities Rulemaking Board, 2022). This ownership concentration amplifies valuation stickiness and redemption risk during stress episodes. As with other segmented markets, these dislocations pose risks to long-term savings portfolios and reveal vulnerabilities in how liquidity is transformed within retail-held asset classes.

Corporate bond ETFs exhibit dislocations of a different nature, more transient but highly reactive to stress. These funds offer daily liquidity but hold underlying assets that trade infrequently in OTC venues. Figure 3b highlights pricing deviations that emerge sharply in episodes like March 2020. The mismatch between ETF share liquidity and underlying bond liquidity becomes pronounced during redemption events, when mechanical selling and valuation opacity widen NAV discounts. These patterns reflect frictions in ETF–OTC intermediation, including the costs and timing of asset transfers across platforms.[9]

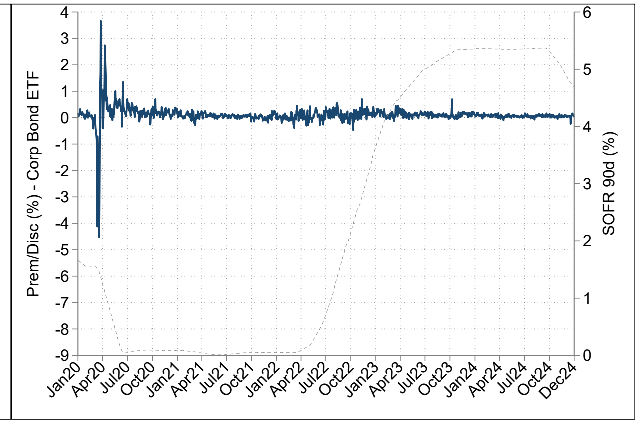
Similar behaviors appear in ETFs holding ABS and other non-government exposures. While NAV deviations in these segments are less extreme, they display consistent sensitivity to funding conditions. These ETFs often hold tranches of securitized assets such as MBS or CLOs, which are subject to valuation opacity and regulatory segmentation. As shown in Figures 3c and 3d, deviations widen during periods of market-wide illiquidity. In particular, CLO ETFs have expanded rapidly since 2023, adding structural complexity to the ETF wrapper. Appendix C provides further descriptive evidence on their growth. The evidence aligns with the findings of He and Song (2025), who document how regulatory segmentation and safe-asset demand generate valuation frictions in agency MBS. ETFs that repackage such assets do not eliminate these frictions but change how they manifest under stress.

These deviations are not isolated anomalies. As emphasized by the European Central Bank (ECB, 2018), deterioration in market liquidity can disable ETF arbitrage channels, leading to persistent NAV dislocations. This fragility is particularly acute in opaque or illiquid segments, where price discovery depends on OTC quotes and dealer willingness to interme-

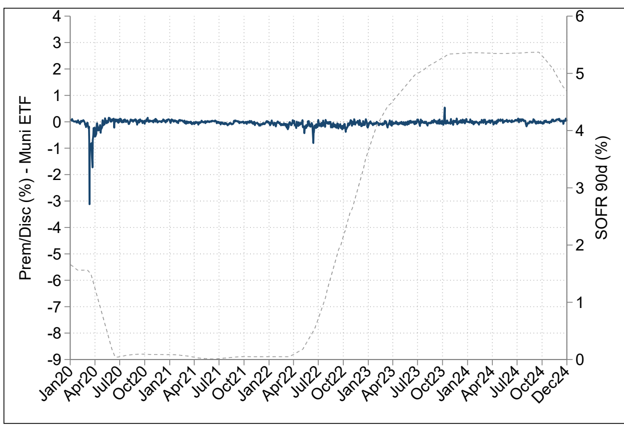
[9]For broader evidence that OTC intermediation costs account for a significant share of credit spread variation during stress periods, see Friewald and Nagler (2019).



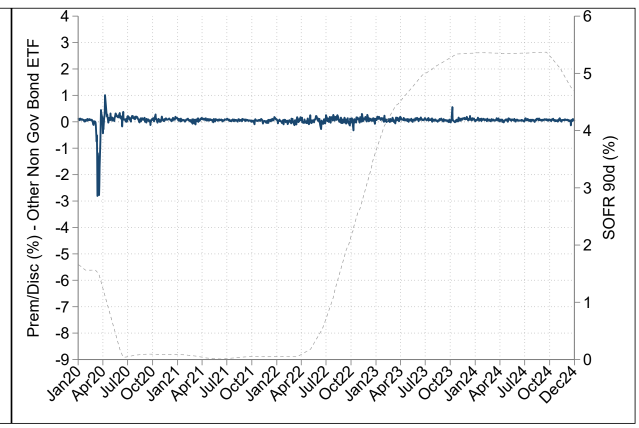
(a) Muni ETF



(b) Corp Bond ETF



(c) ABS ETF



(d) Other Fixed-Income ETF

Figure 3: Premium-Discount and 90-Day SOFR Across Bond ETF Categories, 2020–2024. This figure presents a 2×2 panel of daily ETF premium or discount (blue line, left axis) alongside the 90-day SOFR (gray dashed line, right axis) for four categories of bond ETFs: (a) Municipal Bond ETFs, (b) Corporate Bond ETFs, (c) ABS ETFs, and (d) Other Fixed-Income ETFs. Pricing data are aggregated using market-cap-weighted averages within each category. The sample spans the period from January 2020 through December 2024.

diate. Figure 3 illustrates this breakdown across non-government ETFs. The asymmetry between these dislocated categories and the tight tracking of government ETFs highlights the conditional nature of arbitrage effectiveness.

While several models explain aspects of ETF pricing and liquidity, they fall short of capturing the dynamic interaction between platforms and the endogenous breakdown of arbitrage. Dugast et al. (2022) show how the migration of traders to centralized venues compresses intermediation rents in OTC markets and alters the allocation of risk-bearing. However, these papers do not endogenize ETF share creation or redemption, nor do they model the dual-platform clearing process that drives segmentation in our framework. In the ETF setting, observed share-level liquidity can mask underlying frictions and distorted incentives. Liquidity transformation is not passive: it alters flow dynamics, redemption triggers, and systemic exposures.

Goldstein et al. (2017) document how mutual fund redemptions with illiquid portfolios induce fire-sale dynamics. Unlike mutual funds, ETFs rely on in-kind redemptions and limited cash buffers, transmitting stress through arbitrage pathways rather than cash hoarding—a mechanism our model embeds structurally. They offer liquidity by design but remain exposed to fragmentation in underlying markets.

Hendershott et al. (2022) show how infrequent rebalancing can sustain persistent pricing errors, even in the absence of fundamental shocks. While their model is applied to equities, the underlying logic of participation asymmetries generalizes. Our framework extends this logic to dual trading venues, showing how segmentation and redemption frictions jointly generate persistent ETF–NAV misalignments.

Together, these patterns motivate the segmentation-based framework developed in the next section. They show that fixed-income ETFs operate in a two-speed architecture. ETF shares trade continuously, while their underlying portfolios often reside in illiquid, opaque venues. As arbitrage breaks down during stress, ETF prices diverge from NAV, not because of sentiment or noise, but because of structural frictions that intensify under pressure. The model that follows formalizes this logic and provides a unified lens for interpreting cross-category ETF behavior.

3 Model

Two sets of agents populate our economy. Investors hold on to or trade assets according to both their relative eagerness and ability to trade. Intermediaries facilitate trade while extracting some of the surplus they help generate. Intermediaries also have the skills to transform assets to position them on one or the other platform.

Trees, fruits, and preferences

The environment is composed of a unit measure of investors who derive utility from dividends paid by two types of assets. For concreteness, one can think of the assets as two varieties of trees, each producing their respective type of fruits. The environment has a fixed total number of trees. As detailed below, intermediaries have the skills to 'hybrid' trees from either type to the other and hence determine the number of trees of each type.

The two types of trees are either type-a or type-b[10], and they deliver a single unit of differentiated non-storable fruit each period. These trees differ in the platform on which they get traded: type-a trees require OTC meetings to trade, while type-b can trade at any point in time on exchange.

The exchange venue aggregates order flow and supplies immediacy at a cost. In this dimension, the CEX inherits the logic of Kyle (1985), where order flow conveys information and depth disciplines price impact, and Grossman and Miller (1988), where intermediaries supply immediacy by absorbing temporary order imbalances and bearing inventory risk. Relative to decentralized venue structures that can generate price dispersion (Malamud and Rostek, 2017), the ETF venue in our setting acts as a centralized aggregator that disciplines OTC pricing through creations and redemptions.

Investors derive a utility flow from holding a portfolio (a,b) that produces each time-period a quantity a of type-a fruits, and a quantity b of type-b. Assets are divisible, such that both a and b are non-integer numbers.

Investors are heterogeneous as they get hit by idiosyncratic preference shocks. Each investor utility flow is denoted by $u_i(a,b)$ where i indicates his preference state. There is a finite set of possible preference states denoted by $i \in \mathscr{I} = \{1,...,I\}$. During each period, preference draws arrive at Poisson rate δ, and conditional on their realization, a preference state is drawn identically and independently from the set $\mathscr{I}$ according to the probability distribution $\{\pi_1,...,\pi_I\}$. This time evolution of preferences for holding assets is the motive for trade amongst investors.

Justifications for using a utility framework are discussed in details in the literature[11]. The three most common interpretations are that the utility could (i) come from enjoying the possession of the asset itself, imagine for instance durable assets such as houses or cars; (ii) capture various possible reasons why an investor would like to hold varying quantities of an

[10] In the context of our application, type-a assets represent illiquid securities traded OTC, such as corporate or municipal bonds. Type-b assets correspond to ETF shares backed by these instruments. The model abstracts the liquidity transformation function of ETFs, which offer exchange-traded access to portfolios of otherwise illiquid fixed-income claims.

[11] See Lagos and Rocheteau (2009), footnote 4, on interpreting utility from asset holdings as arising from consumption, liquidity services, or hedging motives—justifying reduced-form specifications in OTC models.

asset, say for instance liquidity or hedging needs[12]; or (iii) be re-interpreted as a production function that takes the two assets as physical capital inputs and produce a unique consumption good.

Regularity conditions are imposed upon the utility function. It is assumed twice continuously differentiable, strictly increasing in all its arguments [13] and strictly concave[14]. Beyond assets, investors derive extra linear utility from consuming a numéraire $c \in \mathbb{R}$ ($c < 0$ is possible, while it is assumed bounded). The quasilinear utility form washes out budget constraints and guarantees the equilibrium uniqueness.

Trading platforms

In terms of trading modalities, on the one hand, type-a assets trade only at OTC meetings. The OTC platform is populated with a unit of intermediaries who mediate all exchanges. Investors can contact the platform after going through a search process. Search frictions encountered to reach intermediaries can be understood as a metaphor for various hindrances. One can think, for instance, of the time required for investors to identify trading partners ready to answer their specific needs, for intermediaries to build tailored contracts, or eventually to conduct a screening process to verify assets' quality and contractors' reliability.[15]

In terms of trading modalities, on the one hand, type-a assets trade only at OTC meetings. The OTC platform is populated with a unit of intermediaries who mediate all exchanges. Investors can contact the platform after going through a search process. Search frictions encountered to reach intermediaries can be understood as a metaphor for various hindrances. One can think for instance of the time required for investors to identify trading partners ready to answer their specific needs, for intermediaries to build tailored contracts, or eventually to conduct a screening process to verify assets' quality and contractors' reliability.

Meetings take place at Poisson rate α. At a realized contact, an investor bargains with the intermediary he has reached. we assume that a Nash bargaining takes place, with bargaining power $\eta \in [0,1]$ for intermediaries (and hence $1-\eta$ to investors). For investors asset price p_a is given, and bargaining takes place upon quantities traded and fees that intermediaries receive per unit traded. Quantities and fees consequently depend upon investors' individual characteristics when trade occurs.

[12] Duffie et al. (2007) liquidity and hedging motives are modeled as preference shocks that shift the marginal utility of holding the asset.

[13] For all $i \in \mathscr{I}$, (a,b) finite, the gradient $\nabla u_i(a,b) >> 0$

[14] From strict concavity, $u_i(.)$ admits a negative semi-definite Hessian matrix $Hess[u_i]$. Hence for all (a,b), $\partial_{11} u_i(a,b) < 0, \partial_{22} u_i(a,b) < 0$, and the Hessian matrix determinant $H[u_i](a,b) > 0$.

[15] The OTC platform refers to decentralized trading mechanisms commonly used for illiquid financial instruments, such as fixed-income securities, structured credit, or bespoke contracts. These environments are characterized by search, bilateral negotiation, and informational frictions. The CEX platform represents centralized venues for continuous trading of liquid claims, such as ETF shares, which transform access to underlying illiquid instruments.

To complete the description, price p_a is set on a competitive market that intermediaries create amongst themselves. This market is assumed to be accessible for nobody else but intermediaries. This inter-intermediaries market represents the netting mechanism across intermediaries who are aiming at a zero residual exposure[16]. Intermediaries are fundamentally market makers whose function is not to hold directional positions, and hence hedge as much as possible of their book holdings.

On the other hand, the exchange (X) platform is catered to trading only type-b assets. It is accessible to investors at all times and performs competitively. The balance of supply and demand for assets b determines its price, given the total quantity quantity of type-b assets to trade upon.

Naturally, within a stationary setting, agents perform trades on this platform in two instances: coincidentally with OTC meetings, and whenever an investor faces an idiosyncratic preference shock. Our model is solvable in closed form because of the opportunity to trade upon both assets when OTC meetings occur.

Assets transformation and transfer

From a market structure perspective, we assume that the total supply Q of assets in the economy is fixed. we denote by $Q^a(t)$ the quantity of a-type asset. And consequently, $Q^b(t) = Q - Q^a(t)$ is the size of the exchange-based market.

Intermediaries have continuous access to both the exchange market and the inter-intermediary markets so that they decide endogenously for the quantity of assets on each platform. They incur a direction-specific cost to transfer an asset from one platform to the other.[17] There is a Υ unit cost to adapt an asset from its OTC form to the exchange's, and Γ is the unit cost to unload an exchange asset onto the OTC. The transfer costs are bounded and potentially outstanding-level dependent.[18]

The structure of the economy is summarized in Figure (4).

3.1 Exchange market

Investors have access to the CEX market at any point in time. They decide how much liquid asset to hold based on the flow utility they derive from their portfolio, moderated by its holding cost. All decisions made on the CEX platform are conditional on the sticky portion of the

[16] A similar modeling logic is used in various strands of literature, see for instance Evans and Lyons (2002)

[17] This mechanism abstracts the ETF creation and redemption process, in which APs convert baskets of illiquid OTC-traded securities into ETF shares, and vice versa. These transfers play a central role in the liquidity transformation function of ETFs, enabling investors to trade illiquid exposures on a continuous CEX platform.

[18] We may assume $\Upsilon'(Q^b) > 0$ to capture the increasing difficulty of creating new products, and $\Gamma'(Q^b) < 0$ to capture the moral hazard issues that come with redeeming complex instruments from a contracting exchange market.

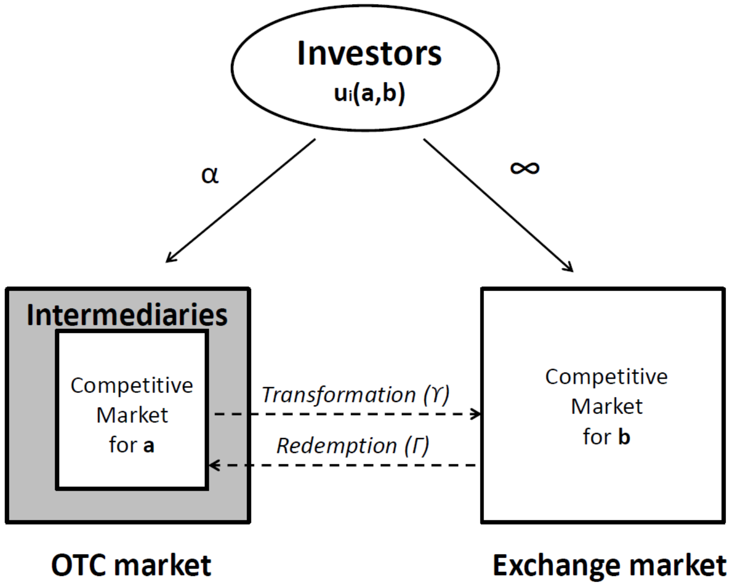


Figure 4: Summary of the model structure with direction-specific transfer costs. Investors contact the OTC platform at rate α to trade asset a, while they can always trade asset b on the CEX platform (contact rate ∞). Transfers from OTC to CEX cost Υ per unit of asset, while the opposite redemption from CEX to OTC costs Γ per unit.

portfolio. Because of the continuous-time setting, the optimization program that determines liquid holdings is identical whether at OTC meetings or in between meetings.

Optimal allocations

A constant risk-free rate r prevails in the economy. At time t, given a continuously differentiable price path $p_b(t)$, the cost to hold one unit of asset b for an infinitesimal period of time is

$$q_b(t) \equiv r p_b(t) - \dot{p}_b(t). \tag{1}$$

This instantaneous holding cost[19] captures the opportunity cost $r p_b(t)$ of fixing one's wealth into the asset instead of a risk-free placement, and is moderated by the per-unit capital gains $\dot{p}_b(t)$.

We define by $b_i(a,t)$ the demand for the liquid asset at time t by an investor with preference type i and holding an illiquid asset position a. From the flow utility function and the holding cost definition

$$b_i(a,t) = \arg\max_{b' \geq 0} \left[u_i(a,b') - q_b(t) b' \right], \tag{3}$$

[19]The reciprocal definition of p_b is immediate. Assuming that the no-bubble condition $\lim_{s\to\infty} e^{-rs} p_b(s) = 0$ is verified,

$$p_b(t) = \int_0^\infty e^{-rs} q_b(t+s) ds. \tag{2}$$

where the time dependency stems from the possible holding costs time-evolution, and hence fades away in the stationary long run. Stated formally,

Lemma 1. *Considering $a \in \mathbb{R}_+$, if $q_b(t) > 0$, then for all $i \in \mathscr{I}$, $b_i(a,t)$ exists and is unique. In particular, $b_i(a,t)$ satisfies the first-order condition*

$$\partial_2 u_i(a, b_i(a,t)) \leq q_b(t) \text{ with equality if } b_i(a,t) > 0, \tag{4}$$

and the transversality condition

$$\lim_{t\to\infty} e^{-rt} p_b(t) b_{i(t)}(a,t) = 0. \tag{5}$$

The existence and uniqueness of the optimal response function are guaranteed by the strict concavity of the utility function with respect to its second argument and the positive price.

Net flow utility and rental market

Since the liquid asset can be traded at any point in time, the optimal conditional holding of type-b assets can be mapped into a typical rental market. The exchange-net flow-utility

$$\widehat{u}_i(a,t) \equiv u_i(a, b_i(a,t)) - q_b(t) b_i(a,t) \tag{6}$$

embeds the liquid margin continuous adjustments. In particular, the utility function $\widehat{u}_i(a,t)$ is now time-dependent from the role of $q_b(t)$, both in the definition of the conditional allocations $b_i(a,t)$ and in the netting itself.

Under the previously stated conditions, the Envelope theorem explicits the marginal benefit of a change in a upon the utility flow when b adjusts continuously

$$\frac{\partial}{\partial a} \widehat{u}_i(a,t) = \partial_1 u_i(a, b_i(a,t)). \tag{7}$$

This technical result ensures closed-form solutions to the OTC platform optimization in the next section.

Illiquidity hedging

The level of substitutability between the two assets in the investors' utility function dictates the cross-impact from illiquid holdings to liquid choices.

Corollary 2. *For interior solutions of optimization (3), the conditional liquid asset demand* *is characterized*[20] *by the substitutability across assets*

$$sign\left(\frac{\partial}{\partial a}b_i(a,t)\right) = sign\left(\partial_{12}u_i(a,b_i(a,t))\right), \tag{8}$$

and negative price elasticity of demand.

The proof is immediate from the implicit function theorem, which states that if $\partial_{22}u_i(a,b_i(a)) \neq 0$, for an interior solution $b_i(a)$,

$$\frac{\partial}{\partial a}b_i(a) = -\frac{\partial_{12}u_i(a,b_i(a))}{\partial_{22}u_i(a,b_i(a))}, \text{ and } \frac{\partial}{\partial q_b}b_i(a) = \frac{1}{\partial_{22}u_i(a,b_i(a))}.$$

Conclusion follows by reminding that the utility function is strictly concave in both its arguments.

In order to simplify the interpretation of later derivations, we assume that for all investors, the substitutability across assets is state-independent. Namely:

Condition 3. *Given* $\zeta \in \{-1,1\}$, $sign(\partial_{12}u_i) = \zeta \quad \forall i \in \mathscr{I}$.

In most economically relevant situations, assets are substitutable to a certain extent, such that in the remainder of this paper $\zeta < 0$.

3.2 OTC market

Investors are able to access the OTC platform after a search period and then meet intermediaries with whom they negotiate the terms of their trade. During a bargaining round, the illiquid asset price is given from the inter-intermediaries market to both investors and intermediaries. Bargaining decides for quantities traded and the fees received by intermediaries.

A Nash bargaining game allows intermediaries to capture a fraction η of the investors' trading gains. This result, derived in Lagos and Rocheteau (2009), greatly simplifies solving the model since all search-and-bargaining trading frictions are captured by the pseudo arrival rate $\kappa = \alpha(1-\eta)$. In particular, the model can equivalently be described as an environment where investors access the inter-intermediaries competitive market for asset a at Poisson rate κ. We denote with τ the corresponding stopping time.

Optimal allocations

The optimization logic for an investor at an OTC meeting is to look forward at the utility flows he expects to derive from his position in the illiquid asset until re-accessing the OTC

[20] For all real x the sign function is defined by $sign(x) = I_{x\geq 0} - I_{x<0}$.

trading platform. To perform this calculation, an investor will consider, on the one hand, the discounted net-utility gains - net of liquid margin adjustments - and on the other hand, the expected holding cost for the illiquid asset.

The holding cost of the illiquid asset a,

$$q_a(t) \equiv (r+\kappa)\left[p_a(t) - \kappa \int_0^\infty e^{-(r+\kappa)s} p_a(t+s)ds\right], \tag{9}$$

corresponds to the current price net of the expected discounted[21] resale price at the next trade opportunity[22].

We define $\widehat{U}_j(a,t)$ the modified utility as the expected net utility flow from current time t to the expected next meeting, such that

$$\widehat{U}_j(a,t) = (r+\kappa)\mathbb{E}_t\left[\int_t^\tau e^{-r(s-t)} \widehat{u}_{i(s)}(a,s)ds \,|\, i(t) = j\right]. \tag{11}$$

In particular the strict concavity of $\widehat{U}$ is implied[23] from the strict concavity of u, the Envelope condition (7), and the implicit function theorem in (3.1).

Hence, an investor with preference type j who contacts[24] a dealer at time t readjusts his a-type asset position by choosing

$$a_j(t) = \arg\max_{a' \geq 0} \left[\widehat{U}_j(a',t) - q_a(t)a'\right] \tag{12}$$

Existence conditions for the optimal demand are given in the following lemma.

Lemma 4. *For any $j \in \mathscr{I}$, if $q_a(t) > \partial_a \widehat{U}_j(\infty,t)$, $a_j(t)$ exists and is unique. Then, $a_j(t)$ satisfies the first order condition*

$$\partial_a \widehat{U}_j(a_j(t),t) \leq q_a(t) \textit{ with equality if } a_j(t) > 0, \tag{13}$$

[21] With τ_m the next meeting time past time t

$$\mathbb{E}_t\left[e^{-r(\tau_m - t)} p_a(\tau_m)\right] = \kappa \int_0^\infty e^{-(r+\kappa)s} p_a(t+s)ds$$

[22] And reciprocally, consistent with Lemma 2 in Lagos and Rocheteau (2009), for any continuous and bounded holding cost $q_a(t)$, the price of an a-type asset is

$$p_a(t) = \frac{1}{r+\kappa}\left[q_a(t) + \kappa \int_0^\infty e^{-rs} q_a(t+s)ds\right]. \tag{10}$$

[23] See Appendix D for details.

[24] For exposition clarity purposes, we follow the convention to denote by j the investor type when negotiating on the OTC market, while denoting by i the investor type when trading the exchange market.

and the transversality condition

$$\lim_{t\to\infty}\mathbb{E}\left[e^{-rT_t}p_a(T_t)a_{i(T_t)}(T_t)\right]=0, \tag{14}$$

where T_t denotes the last contact with an intermediary before time t.

The objective function on the right-hand-side of the maximization (12) is strictly concave and differentiable, such that the first order condition (13) is necessary and sufficient for an optimum. Furthermore, from the price restriction, the solution is unique.

Intermediaries

Intermediaries are entitled to capture a fraction η of the trade gains from OTC meetings, and consequently, as shown in the appendix, the fees they capture at time t from trading with a type-i investor who holds inventory a are

$$\phi_i(a,t)=\eta\left[\frac{\widehat{U}_i(a_i(t),t)-\widehat{U}_i(a,t)}{r+\kappa}-\frac{q_a(t)}{r+\kappa}[a_i(t)-a]\right]. \tag{15}$$

At the aggregate level, during a period of time dt, a fraction αdt of the population contacts an intermediary to perform OTC trades. Assuming that we know $M_i(a,t)$, the measure of investors at time t who are of preference type i and holding a position a in the illiquid asset, from the law of large numbers, intermediaries capture an aggregate amount of fees

$$\Phi(t)=\alpha\sum_{i\in\mathscr{I}}\int_0^{\infty}M_i(a,t)\phi_i(a,t)da. \tag{16}$$

3.3 Demand aggregation

We first detail the distribution of agents in the economy at every point in time, and then derive aggregate demand functions on each platform.

Measure of investors

From the idiosyncratic preference shocks structure, the measure of individuals in state i at time t is $n_i(t)=e^{-\delta t}n_i(0)+(1-e^{-\delta t})\pi_i$. And the probability for an investor to transition from preference-state k to preference-state i during the lapse of time Δ is $p_{k,i}(\Delta)=(1-e^{-\delta\Delta})\pi_i+e^{-\delta\Delta}I_{k=i}$. Hence, by Lemma 3 in Lagos and Rocheteau (2009) , the measure of agents holding any illiquid position $a\in\mathbf{a}$ –where $\mathbf{a}$ is a set of possible holdings– and in state i at time t is

$$M_i(\mathbf{a},t)\equiv m_i^0(\mathbf{a},t)+\int_0^t m_i(\mathbf{a},t,\tau)d\tau. \tag{17}$$

The first term captures the measure of investors who held the target quantity of illiquid assets at time zero, and have not gone through any OTC meeting since then. Since the probability of not undergoing any OTC meeting during a lapse of time Δ is $e^{-\alpha\Delta}$, their measure is exactly

$$m_i^0(\mathbf{a},t) = e^{-\alpha t} \sum_{k\in\mathscr{I}} p_{k,i}(t) M_k(\mathbf{a},0). \tag{18}$$

The second term considers the measure of investors who have acquired the target illiquid asset quantity at their last OTC meeting, which meeting may have occurred at any point in time. Given $\alpha e^{-\alpha\tau}$ the probability to have reached an intermediary τ time-periods ago, the measure of investors who have last traded to reach the target illiquid holding level is

$$m_i(\mathbf{a},t,\tau) = \alpha e^{-\alpha\tau} \sum_{k\in\mathscr{I}} p_{k,i}(\tau) n_k(t-\tau) I_{\{a_k(t-\tau)\in\mathbf{a}\}}. \tag{19}$$

Aggregate demand functions

Investors, at the CEX platform, ask for the optimal holding relative to their current preference state and illiquid asset holding, so that the aggregate demand

$$D^b(t) = \sum_{i\in\mathscr{I}} \int_0^\infty M_i(a,t) b_i(a,t) da. \tag{20}$$

On the OTC platform, making use of the law of large numbers[25], over a time period dt, the demand of asset a is equal to $\alpha \sum_{j\in\mathscr{I}} n_j(t) a_j(t) dt$. This latter formula comes from the aggregation of demand for the illiquid asset, across agents of any possible preference type. Using a shorthand notation, the potential demand is

$$D^a(t) = \sum_{j\in\mathscr{I}} n_j(t) a_j(t). \tag{21}$$

3.4 Adjustment of market sizes with active role for intermediaries

Conditional on prices, optimal aggregate demand for assets is as developed earlier. Hence, as the total of all assets is fixed, transfers are set in equilibrium by the balancing of marginal cost and marginal benefit.

What about the fees? The choice of market depth for the OTC segment has an impact on the fees intermediaries will be able to capture going forward. This tension is captured through the price mechanism in the inter-intermediaries market. At every point in time, an

[25] Formally, this aggregation relies on the existence of a continuum random matching process with an exact law of large numbers, as established by Duffie and Sun (2007).

intermediary who has been contacted by an investor is indifferent between obtaining the asset in the inter-intermediaries market, and 'importing' it himself from the exchange market. The price-taking structure prevents arbitrage opportunities from arising.

At any point in time, all investors can trade on the exchange, so that supply and demand instantaneously clear as

$$D^b(t) = Q^b(t) = Q - Q^a(t), \tag{22}$$

and determine the trading price $p_b(t)$.

On the OTC platform, intermediaries can modify the availability of OTC assets in response to investors' demand. Given the contact rate α, the corresponding continuous adjustment flows are

$$\dot{Q}^a(t) = \alpha \left[D^a(t) - Q^a(t)\right]. \tag{23}$$

Detailing the potential[26] demand function $D^a(t)$, the measure $n_i(t)$, does not depend on any contemporaneous variable level, while $a_j(t)$ depends negatively on holding cost $q_a(t)$ and positively on holding cost $q_b(t)$. Hence at a stationary equilibrium for which $\dot{Q}^a(t) = 0, ie.$ $Q^a(t) = D^a(t)$, the demand for asset a is negatively correlated to $p_a(t) - p_b(t)$. Conversely, since $\dot{Q}^b(t) = -\dot{Q}^a(t)$, at equilibrium $Q^b(t) = D^b(t)$ and the demand for asset b is negatively correlated to $p_b(t) - p_a(t)$.

The objective function of an intermediary is to maximize his wealth W that comprises the fees from OTC trades, and the proceeds from readjusting the exchange market depth Q^b. The wealth of an intermediary at time t while the current quantity of exchange asset is $Q^b(t) = Q^b$ is

$$W(Q^b, t) = \int_t^\infty e^{-r(s-t)}\Phi(s)ds + \max_{\{Q^b(s)\geq 0\}} \int_t^\infty e^{-r(s-t)}\Psi\left(Q^b(s), s\right) ds. \tag{24}$$

The first term is the time-t expected discounted fees captured by intermediaries from time t onward. From the definition of fees in equation (16), this function is independent of the size of the OTC market, which impacts fees only through the price mechanism.

The second term is the time-t discounted gain from transforming and transferring assets

$$\Psi\left(Q^b(s), s\right) = \left[p_b(s) - p_a(s) - \Upsilon\left(Q^b(s)\right) I_{\dot{Q}^b(s)>0} + \Gamma\left(Q^b(s)\right) I_{\dot{Q}^b(s)<0}\right] \dot{Q}^b(s). \tag{25}$$

In detail, to increase the exchange market pool by one unit, a representative intermediary buys one unit of the asset on the OTC inter-intermediaries for p_a, incurs the unit cost Υ, and resells the asset for p_b on the exchange market. Conversely, to transfer back an asset to the

[26] Only the fraction of agents who get to meet an intermediary get to express this demand.

OTC market entails buying it at price p_b, incurring the unit redemption cost Γ, and reselling the asset on the OTC market for price p_a.

By optimization, $\dot{Q}^b(s)$ is positive or null when $p_b(s) - p_a(s) \geq \Upsilon\left(Q^b(s)\right)$, and negative or null when $p_a(s) - p_b(s) \geq \Gamma\left(Q^b(s)\right)$. Whenever the price differential $p_b(s) - p_a(s)$ belongs to the open set $\left(-\Gamma\left(Q^b(s)\right), \Upsilon\left(Q^b(s)\right)\right)$ then $\dot{Q}^b(s) = 0$ which also means that the demand for asset-b is satisfied with the assets on the centralized platform, i.e., $Q^b(s) = D^b(s)$.

Lemma 5. *Suppose that the price paths $p_a(s)$ and $p_b(s)$ are differentiable, and that the initial condition $Q^b(t) = Q^b$ is such that $p_b(t) - p_a(t) \notin \left[-\Gamma\left(Q^b\right), \Upsilon\left(Q^b\right)\right]$. Then, a bounded path, $Q^b(s)$ solves the intermediaries' problem, if and only if it verifies the no-bubble condition*

$$\lim_{T \to +\infty} e^{-rT} \left[p_b(T) - p_a(T)\right] Q^b(T) = 0, \tag{26}$$

together with the optimality condition for all $s > t$

$$r\left[p_b(s) - p_a(s) - \Upsilon\left(Q^b(s)\right) I_{\dot{Q}^b(s)>0} + \Gamma\left(Q^b(s)\right) I_{\dot{Q}^a(s)<0}\right] = \Xi(s) \tag{27}$$

whenever $\dot{Q}^b(s) \neq 0$; and where the dynamic part $\Xi(s)$ is defined by

$$\Xi(s) = \dot{p}_b(s) - \dot{p}_a(s) - \dot{Q}^b(s)\left[I_{\dot{Q}^b(s)>0} \Upsilon'\left(Q^b(s)\right) - I_{\dot{Q}^b(s)<0} \Gamma'\left(Q^b(s)\right)\right]. \tag{28}$$

The proof of this lemma proceeds by integration by part of the objective function and is detailed in Appendix G.

The phase diagram representative of the economy in Figure 5 has a three-zone structure in the $(Q^b, p_b - p_a)$ plane. The intermediate zone sees no adjustment in the quantity of assets between platforms.

3.5 Equilibrium

We first define the equilibrium across holdings and prices, so as to then state the existence of an equilibrium in proposition (7).

Definition 6. *An equilibrium is a time path of allocations, prices and distribution*

$$\left\langle \left\{a_j(t)\right\}, \{b_i(a,t)\}, D^a(t), D^b(t), q_a(t), p_a(t), q_b(t), p_b(t), \{\phi_j(a,t)\}, Q^a(t), \left\{M_i\left(a_j(t), t\right)\right\} \right\rangle$$

that satisfies for all $(i,j) \in \mathscr{I}^2$ (4),(13),(20), (21), (9), (1), (22), (23), (15), (27), and (17), given the total Q and an initial condition M^0.

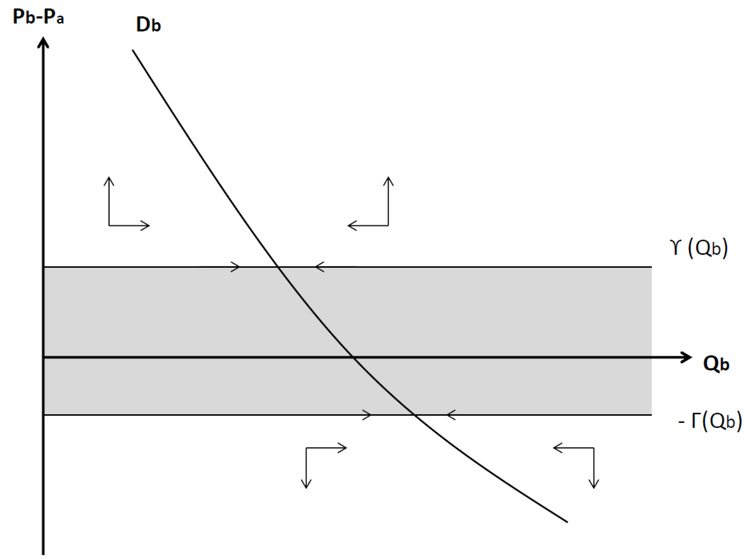


Figure 5: Phase Diagram with direction-specific transfer costs. On the y-axis the differential in prices $p_b - p_a$, and on the x-axis the outstanding size of the b-asset platform. The transformation $\Upsilon(Q^b)$ and redemption $-\Gamma(Q^b)$ frontiers are plotted together with a representative asset-b demand curve D^b.

In Appendix F, we follow the logic of exercise 10.G.1 in Mas-Colell et al. (1995) , adapting for the fact that the OTC market is not Walrasian to show that

Proposition 7. *for any initial type and illiquid asset distribution M^0 in the population there exists a unique equilibrium.*

Deriving the time-t equilibrium result hinges on the property of the agents' utility functions. All agents extract from the three goods a utility that is quasilinear with respect to the same numéraire good, which rules out wealth effects, and leads the market for numéraire to clear automatically. Our novel proof adapts the typical planner proof in frictionless markets by assuming that the planner incorporates both OTC market frictions (search and bargaining) when allocating optimally assets amongst agents.

Stationary equilibrium

Convergence of the previous equilibrium to its stationary counterpart is described in the lemma that follows. Before stating the technical lemma, stationary net flow utility is defined for all $i \in \mathcal{I}$ and possible illiquid asset holding a as $\widehat{u}_i(a) = u_i(a, b_i(a)) - q_b b_i(a)$, and the transformed utility is for any $j \in \mathcal{I}$

$$\widehat{U}_j(a) = \frac{(r+\kappa)\widehat{u}_j(a) + \delta \sum_i \pi_i \widehat{u}_i(a)}{r+\kappa+\delta}. \tag{29}$$

The transformed utility captures the expected utility derived from holding an illiquid portfolio a while undergoing any possible number of preference shocks and being unable to adjust this margin.

In the stationary long run, illiquid holding choices are solely indexed by the investor's type at the OTC meeting. We denote by $M_{j,i}$ the stationary measure of agents in state i and with illiquid holding a_j, so that for all $(i,j) \in \mathscr{I}^2$ it is simply expressed as $M_{j,i} = \pi_j \left(\delta\pi_i + \alpha I_{i=j}\right)/(\delta+\alpha)$ and by aggregation the measure of investors holding position a_j is $\sum_{i\in\mathscr{I}} M_{j,i} = \pi_j$.

Definition 8. *A steady state of allocations, prices and distributions*

$$\left\langle \{a_j\}, \{b_i(a)\}, D^a, D^b, q_a, p_a, q_b, p_b, \{\phi_j(a)\}, Q^a, M_{j,i} \right\rangle$$

has its components characterized, for all $(i,j) \in \mathscr{I}^2$, *by: (i) the price to holding cost relationships* $q_a = rp_a$ *and* $q_b = rp_b$*; (ii) the first order conditions on the OTC platform* $rp_a = \widehat{U}'_j(a_j)$*, and on the exchange* $rp_b = \partial_2 u_i(a, b_i(a))$*; (iii) the two aggregated demands* $D^a = \sum_{j\in\mathscr{I}} \pi_j a_j$ *and* $D^b = \sum_{(i,j)\in\mathscr{I}^2} M_{j,i}\, b_i(a_j)$*; (iv) the two market clearing conditions* $D^a = Q^a$ *and* $D^b = Q - Q^a$*; (iv) the fees extracted by intermediaries,* $\phi_j(a) = \eta\left[\widehat{U}_j(a_j) - \widehat{U}_j(a) - rp_a\left(a_j - a\right)\right]/(r+\kappa)$*; (v) the equilibrium measure of investors* $M_{j,i} = \pi_j\left(\delta\pi_i + \alpha I_{i=j}\right)/(\delta+\alpha)$ *and (vi) depending on the initial condition* Q_0^b*, one of the three conditions (vi-1)* $Q^b = Q_0^b$*, (vi-2)* $p_b - p_a = \Upsilon$*, or (vi-3)* $p_b - p_a = -\Gamma$*.*

Lemma 9. *An equilibrium of allocations and prices as defined in (6) converges to a steady-state equilibrium of allocations and prices as defined in (8).*

We now turn to a quantitative analysis of the model. To illustrate the mechanics and implications of the dual-platform structure, we simulate portfolio allocations, price dynamics, and equilibrium adjustments under various trading frictions and investor preference shocks. The numerical implementation tracks the evolution of key outcomes under both fixed and endogenous platform sizes, and allows us to assess the effects of liquidity constraints, asset segmentation, and market freezes. The remainder of our exposition focuses on stationary equilibria.

4 Simulation Environment

4.1 Utility and preferences

To put our model to work, we consider the utility specification:

$$u(a,b) = \epsilon \frac{\left[s a^{1-1/\rho} + (1-s)\left(\mu b\right)^{1-1/\rho} \right]^{\frac{1-\gamma}{1-1/\rho}}}{1-\gamma}. \tag{30}$$

Investors are assumed to have constant relative risk aversion (CRRA) defined over a constant elasticity of substitution bundle of the two assets. The aggregator captures constant elasticity ρ between the two assets' flows and a utility share parameter s. The parameter μ captures the factor-specific productivity of the liquid asset, shifting the relative marginal utility of b versus a. A few comments are in order. First, the CRRA formulation ensures that marginal utility is infinite at the origin, which guarantees strictly positive holdings in both assets. Second, preferences feature both curvature over the bundle (through γ, the coefficient of relative risk aversion, which reflects how sensitive investors are to overall fluctuations in consumption) and substitutability within it (through ρ, the elasticity of substitution, which describes how easily investors shift between the two assets), allowing us to capture both risk aversion and flexibility in portfolio adjustment. Lastly, the factor-specific productivity parameter μ captures how much agents relatively value the assets traded on either platform. It serves as an asset-specific valuation channel that reflects perceived differences in liquidity, access, or convenience. In the baseline parametrization, we assume that assets traded on the OTC platform generate a higher utility flow.

4.2 Calibration and baseline scenario

The simulation builds on the utility formulation above which introduces heterogeneity through preference shocks. In particular, it embeds three potential sources of idiosyncratic variation: (i) a productivity shock to ϵ, which affects the marginal value of both assets simultaneously; (ii) a shock to μ, which alters the relative desirability of one asset over the other; and (iii) a shift in the utility share parameter s, which influences the desired asset mix directly.

In our simulations, we focus on the first case (i). Investors switch between two productivity levels, ϵ_H and ϵ_L, with $\epsilon_H > \epsilon_L$, which triggers portfolio rebalancing. From the shock structure, we have $a_H \geq a_L$, $b_H(a_H) \geq b_L(a_H)$, and $b_H(a_L) \geq b_L(a_L)$. The substitutability property further implies that $b_H(a_L) \geq b_H(a_H)$ and $b_L(a_L) \geq b_L(a_H)$.

These inequalities describe how both the level effects (due to higher utility scaling) and substitution effects (across asset holdings) shape the demand response of investors under preference switching. The simulations quantify the size of these responses and provide a benchmark for analyzing portfolio dynamics and liquidity allocation under heterogeneous preferences. This echoes preferred-habitat models, where segmentation and limited arbitrage generate persistent price distortions (Vayanos and Vila, 2021); in our setting, the segmentation is across venues rather than maturities. The baseline parameter values are summarized in Table 1.

Parameters		Value	Description
Markets			
	r	0.1	Discount rate
	α	1	Arrival intensity of meetings
	η	0.1	Intermediaries' bargaining power
	δ	2	Arrival intensity of type-switches
	π	$(0.6, 0.4)$	State probabilities
Utility			
	γ	0.5	Relative risk aversion
	ρ	5	Elasticity of substitution between assets
	ϵ	$(5, 10)$	Productivity shock
	s	$(0.5, 0.5)$	Share parameter
	μ	0.95	b factor specific productivity
Platforms			
	Q	1	Total supply of assets
	Υ	5	Unit cost to engineer an exchange-traded asset
	Γ	5	Unit cost to redeem an exchange-traded asset

Table 1: Baseline parameters for the simulation environment. The calibration is consistent with LR and reflects a stationary environment with preference switching, trading frictions, and platform segmentation.

4.3 Fixed platform sizes and investor portfolio dynamics

We illustrate in Figure 6, under a stationary environment, how an individual investor's portfolio allocation evolves as they experience a sequence of preference shocks and OTC meetings. The preference shocks are identified by the jumps in the ϵ process between two levels (high and low), and the OTC meetings are indicated by stars along the ϵ trajectory. Platform sizes are fixed to half of the total assets in the economy. Solid lines represent allocations in the OTC–CEX environment, while dashed lines represent the OTC-only environment. For illustration, in the OTC-only case both asset types are assumed to trade under OTC rules. Red denotes b-type assets, and blue denotes a-type assets.

The three panels of the figure should be read top to bottom. The top panel displays the sequence of preference shocks, alternating between high and low states. Each time the in-

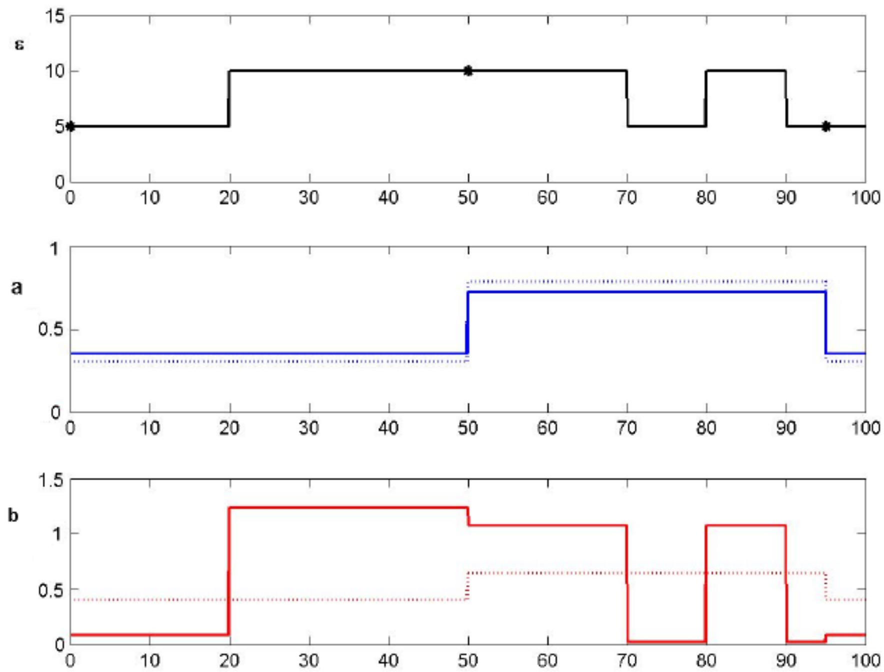


Figure 6: Sample portfolio allocation path for a representative investor. The top chart shows the path of preference states, the middle chart the illiquid asset choice, and the bottom chart the liquid asset allocation. OTC meetings are highlighted with stars in the top chart. Solid and dotted lines represent the OTC–CEX environment and the OTC-only environment, respectively, for each asset on its fixed-size platform.

vestor is hit by a shock, they adjust their portfolio allocation, but only when they meet an intermediary, indicated by a star.

The center panel tracks the allocation to the illiquid asset a. Under the OTC–CEX environment (solid blue line), the investor adjusts more cautiously, leading to smaller jumps in holdings. In contrast, under the OTC-only environment (dotted blue line), portfolio adjustments are less responsive and may deviate more from the investor's desired allocation, as rebalancing is constrained by limited access to trading opportunities.

The bottom panel shows the allocation to the liquid asset b. Here, we observe more pronounced rebalancing in the OTC–CEX environment (solid red line), where the investor takes advantage of immediate trading opportunities on the CEX. By contrast, in the OTC-only environment (dotted red line), the investor is only able to adjust through occasional OTC meetings, leading to delayed and more muted portfolio responses. Note that in the OTC-only environment, both assets a and b are traded through OTC meetings; the red dotted line reflects investor holdings of asset b when CEX access is unavailable.

The evolution of both red lines from left to right should be read as the investor's successive responses to preference shocks. At each meeting (marked by a star in the top panel), the investor updates their allocation to b, subject to current platform constraints. In the OTC–CEX environment, these adjustments can be immediate and sizable; in the OTC-only environment, they are slower and less reactive, as the investor must wait for OTC access to make portfolio changes.

Two immediate observations. First, in the illiquid market, the OTC–CEX scenario features more conservative positions and therefore smaller trade volumes than the OTC-only environment. By contrast, in the liquid market, trade volumes increase substantially. Second, on the CEX platform, investors hold more extreme positions between the time they are hit by a preference shock and the next OTC meeting. This phenomenon, which we refer to as illiquidity hedging, is characterized in Lemma 8.

4.4 Endogenous platform size and equilibrium segmentation

From a state of the economy where all assets are traded on the OTC platform, intermediaries start transferring assets to the exchange market only if the valuation of these assets by investors is sufficient to remunerate the transfers. The stock Q_b grows until balance is struck between demand and supply, and the price difference between assets on either platform equals their per-unit transfer cost, $p_b - p_a = \Upsilon$. In the case of variable creation costs (dependent on platform size or other factors), the build-up of the exchange platform may not always be warranted.

As illustrated in Figure 7, the initial demand curve D_b shifts to the right to D'_b in response to a permanent increase in demand for exchange-traded assets.[27] The nature of the adjustment depends on the economy's initial location along D_b.

If the economy starts at the lower-right point on the lower boundary, the adjustment is vertical: the price differential $p_b - p_a$ rises while the exchange quantity Q_b remains unchanged. If the starting point is at the upper-right corner, already on the upper boundary, the adjustment is primarily horizontal. The price remains near the transfer threshold, and the exchange share Q_b increases to absorb the added demand. For interior starting points, a small enough demand shock leads to a purely vertical move: prices rise within the shaded no-transfer band, but Q_b stays constant. However, if the shock is large enough, the price crosses the transfer boundary and the adjustment becomes horizontal, expanding Q_b and shifting the equilibrium to the right. In that case, the transition combines both a price effect and platform expansion.

[27] Ben-David et al. (2023) document durable segmentation within the ETF venue along fees and product scope. Broad-based, low-fee funds coexist with more specialized, higher-fee funds that display higher turnover and lower risk adjusted performance. This investor sorting is consistent with demand shifts that, once transfer frictions are overcome, expand the exchange platform in our framework.

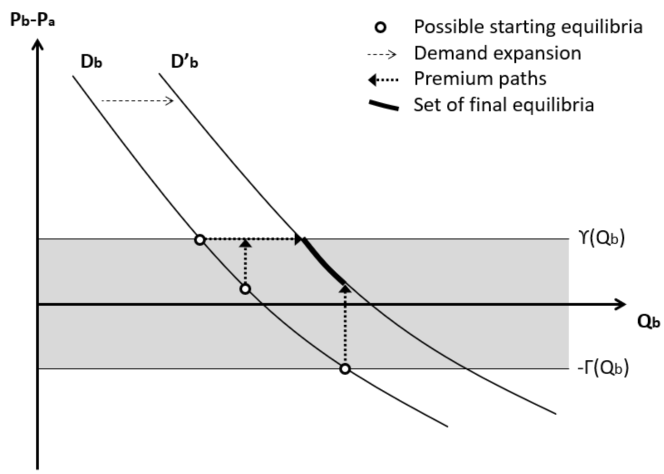


Figure 7: Impact of a permanent demand shock. From the original potential loci of equilibrium—along the D_b curve which lies within the transfer boundaries, i.e., within or at the limit of the greyed area—the diagram schematically illustrates the set of equilibria reached after a permanent increase in demand for exchange-traded assets. The arrows reflect how equilibrium adjusts depending on the initial location: through changes in the price differential, in the exchange share Q_b, or both.

The shaded region between $-\Gamma(Q_b)$ and $\Upsilon(Q_b)$ defines the range in which price differentials are insufficient to trigger reallocation between platforms. Movements within this band occur through price adjustment only. Once the boundary is reached, transfers become active, and the system adjusts through quantity as well. The bold segment on D'_b highlights the final set of attainable equilibria following the demand shock, conditional on the starting configuration. Viewed through the lens of Rahi and Zigrand (2009), this no-transfer band is the region where participation costs deter cross-segment trades, so optimal security design by arbitrageurs does not fully eliminate venue-specific price differences.

These comparative static results uncover the underlying mechanics of fragility and reallocation in dual-platform markets, and show how platform segmentation shapes equilibrium behavior under varying demand conditions. We now turn to dynamic simulations that operationalize these forces and examine how investor preferences and liquidity constraints propagate through the system to generate asymmetric responses, volatility, and systemic fragility.

5 Results

Having developed the theoretical framework, we now explore its implications through comparative statics. The model captures key mechanisms observed in fixed-income markets, including segmentation, investor reallocation, and venue-specific frictions. In the sections that follow, we link these dynamics to stylized facts and simulation results, focusing on how ETF-driven liquidity transformation affects asset pricing, investor welfare, and systemic vulnerability.

5.1 OTC liquidity frictions and demand for exchange option

This section examines how trading frictions in OTC markets shape investor utility and asset allocation across venues. Within our model, OTC illiquidity is proxied by the matching speed parameter α, which governs how frequently investors can adjust their holdings through bilateral negotiation. We simulate comparative statics across a range of α values to quantify the welfare effects and platform demand under two benchmark regimes. In the first, an OTC-only setting, all assets are traded exclusively through decentralized search. In the second, an OTC–CEX environment, both platforms are accessible, and the share of assets traded on each emerges endogenously in equilibrium.

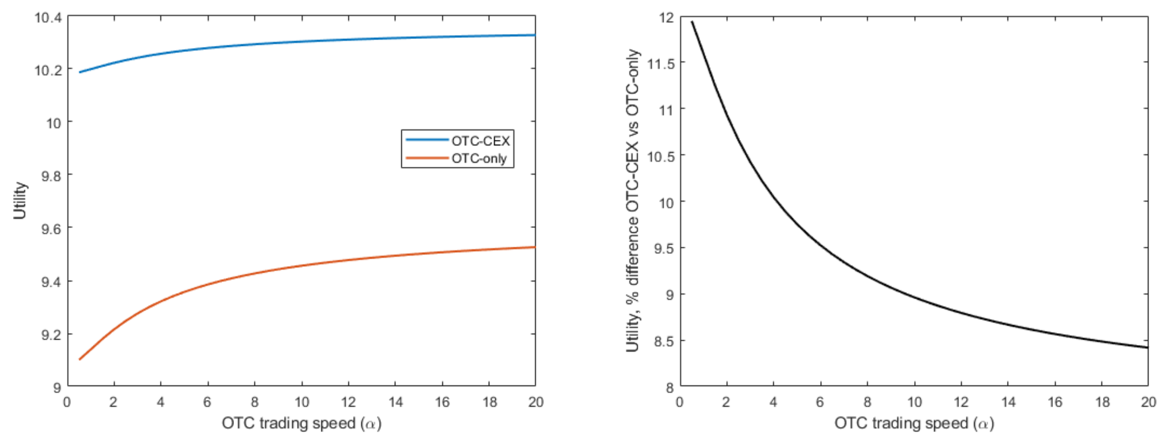


Figure 8: Investor utility in OTC-only and OTC–CEX environments under different OTC trading frictions. On the x-axis, the OTC trading speed α. The left panel shows utility levels in the OTC-only and OTC–CEX environments, while the right panel displays their percentage difference.

As shown in Figure 8, utility in both regimes converges at high values of α, when OTC frictions are minimal. In this region, frequent bilateral meetings allow investors to rebalance efficiently, diminishing the value of access to the CEX. However, as trading frictions increase, i.e., when α decreases, investor utility declines more rapidly in the OTC-only regime. The gap between the two settings widens, reflecting the rising marginal value of liquidity access through the exchange. In effect, the presence of a liquid platform serves as a hedge against rebalancing constraints, with increasing payoff as OTC intermediation slows.

Figure 9 illustrates the corresponding shift in asset allocation. As OTC matching speed

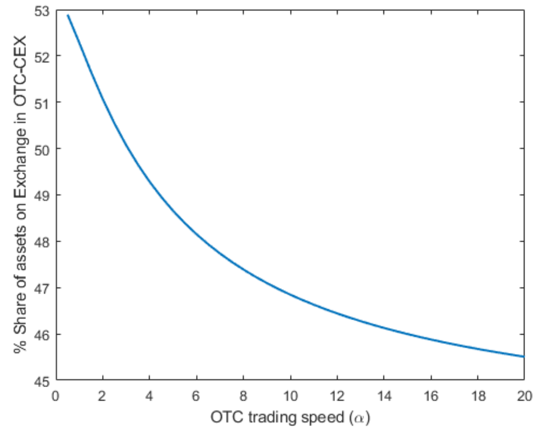


Figure 9: Allocation of assets to the CEX platform in the OTC–CEX environment. On the x-axis, the OTC trading speed α. On the y-axis, the share of assets optimally allocated to the CEX platform.

slows, the optimal share of assets transferred to the CEX rises. When OTC access becomes unreliable, investors endogenously substitute toward the more liquid venue, even when this option offers imperfect exposure. This pattern captures a form of liquidity-seeking behavior under segmentation: investors optimally shift their portfolios to maintain tradability when one platform becomes impaired.

These dynamics are consistent with equilibrium models in which investor inertia or slow rebalancing sustains cross-platform allocations even under persistent pricing differentials. As Hendershott et al. (2022) emphasize, frictions in portfolio adjustment can generate segmentation that endures despite observable arbitrage opportunities. In our setting, the CEX does not eliminate pricing gaps, it provides a mechanism for managing liquidity constraints when OTC trading becomes less viable. Importantly, the relative benefit of the exchange option increases nonlinearly with OTC illiquidity, reinforcing the idea that segmentation magnifies the utility cost of frictions.

In sum, the simulations reveal how the presence of a secondary venue buffers investors against illiquidity risk. The endogenous response, in which assets shift toward the CEX as frictions grow, highlights the demand for flexible liquidity access under segmentation. These results provide the theoretical foundation for the ETF behavior observed in segmented credit markets and set the stage for our analysis of price formation and spread behavior in the next sections.

5.2 ETF creation reducing OTC bid-ask spreads

Building on the segmentation dynamics discussed in Section 5.1, we now examine how the introduction of ETF-based trading affects liquidity in OTC markets. Specifically, we investigate how access to an exchange-traded alternative reshapes investors' willingness to incur intermediation costs and how this transformation endogenously compresses bid-ask spreads.

In our model, liquidity is structurally defined as the ease of contacting a counterparty through OTC matching, captured by the trading speed α. In practice, bid-ask spreads serve as the traditional observable proxy for market liquidity. We estimate this spread analog structurally, using the unit fees captured by intermediaries on both the buy and sell sides of OTC transactions.

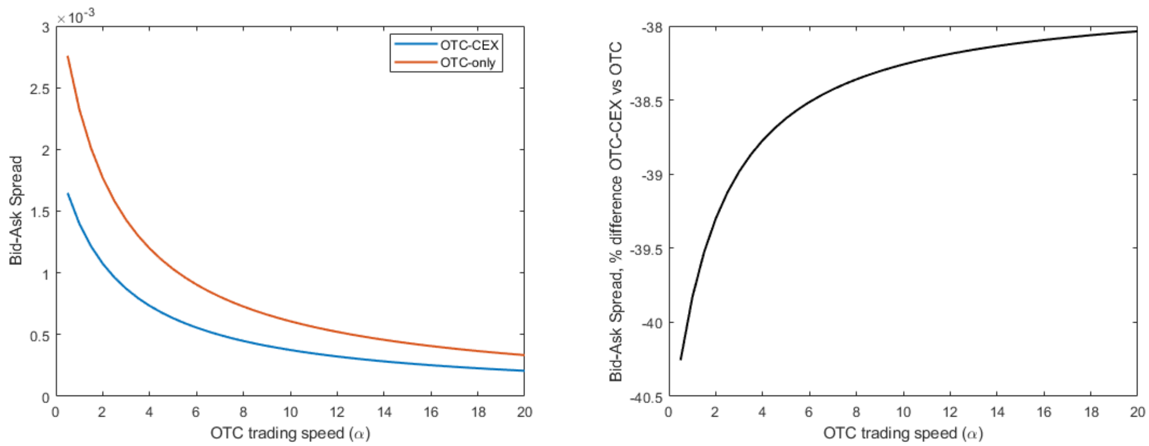


Figure 10: Bid–ask spreads in OTC-only and OTC–CEX environments under different OTC trading frictions. On the x-axis, the OTC trading speed α. The left panel shows bid–ask spreads in the OTC-only and OTC–CEX environments, while the right panel displays their percentage difference.

Figure 10 illustrates how bid-ask spreads respond to OTC trading frictions across two environments. As trading speed decreases (i.e., as frictions grow), intermediation costs rise, resulting in wider spreads. However, when a CEX venue is present, the ability to shift demand toward a continuously accessible market suppresses investors' willingness to pay high OTC fees. As a result, equilibrium spreads in the OTC venue are significantly lower in the OTC–CEX setting.[28]

This effect emerges from a reallocation of liquidity demand. The availability of a liquid substitute on the CEX platform reduces the expected gain from trade in the OTC venue, diminishing the fees intermediaries can charge. Even when investors ultimately choose to

[28]Figure A2 in Appendix details the composition effect in total fees.

remain in OTC markets, the shadow cost of exit disciplines fee-setting behavior.[29]

These dynamics mirror the venue competition framework in Pagnotta and Philippon (2018), in which faster platforms erode the rents of slower ones through investor sorting and price competition. Our results also align with Holden et al. (2021), who show that segmented markets can sustain non-overlapping pricing regimes, with investors migrating toward venues offering more efficient execution. Moreover, the mechanism is conceptually related to Carapella and Monnet (2020), who demonstrate how central clearing compresses intermediation spreads but reduces dealers' ex ante incentives to provide liquidity, a tension that parallels the effect of ETF trading on OTC fee structures.

In sum, when investors have access to liquid ETF shares, the marginal value of OTC execution declines, leading to an endogenous reduction in bid-ask spreads. This compression is not the result of improved OTC functioning per se, but of declining investor tolerance for illiquidity when a continuously tradable alternative exists. The next section examines how these frictions influence ETF price behavior when redemption becomes impaired.

5.3 OTC freeze and ETF price-discovery

The 2007–2009 financial crisis exposed the fragility of OTC derivatives markets, which experienced an abrupt collapse in trading activity. In response, regulators, including the Group of Twenty (G20), advocated for the migration of standardized derivatives to more transparent and liquid venues. This regulatory shift laid the foundation for the expansion of exchange-traded instruments, including fixed-income ETFs, which offer continuous tradability even when underlying markets become impaired.[30]

Within our model, we investigate the consequences of an unanticipated freeze of the OTC platform on price formation in the remaining CEX platform. While traditional models such as Guerrieri and Shimer (2014) emphasize informational frictions as the source of market freezes, our framework highlights a distinct and complementary mechanism: the reallocation of investor demand from the illiquid to the liquid platform in a segmented market. This shift alters the structure of equilibrium by placing upward pressure on prices without requiring changes in beliefs or fundamentals.

This dynamic generates a form of path-dependent unraveling. When decentralized trad-

[29] Coupling of unit fees and volumes is detailed in Appendix H.

[30] Following the 2009 Pittsburgh and 2010 Toronto G20 summits, standardized OTC derivatives were mandated to migrate to exchanges or platforms, be centrally cleared, and be reported to trade repositories, with higher capital charges for uncleared trades. See *G20 Leaders' Statement, Pittsburgh Summit*, September 24–25, 2009; *G20 Toronto Summit Declaration*, June 26–27, 2010. For a theoretical assessment, Duffie and Zhu (2011) show that central clearing can reduce bilateral exposures but also reallocate risks, leaving systemic fragility largely unchanged.

ing halts, the CEX platform absorbs residual liquidity demand and becomes the sole venue for price formation. This transition is consistent with the mechanism described by Rust and Hall (2003), where centralized markets gradually attract order flow as their relative accessibility increases, even in the absence of fundamental shocks. It also echoes the fragility mechanism in Hendershott et al. (2024), where strategic dealer withdrawal can abruptly eliminate liquidity provision in OTC venues, accelerating the migration toward more transparent platforms. Our model embeds this reallocation structurally and traces its implications for ETF pricing, capturing how price pressure intensifies even without changes in beliefs or fundamentals.

We consider a scenario in which trading on the OTC platform is suddenly and entirely suspended. We abstract here from the cause of the disruption. From that point forward, type-a assets become non-tradable, and investors must retain their illiquid holdings. If the freeze occurs in a stationary environment, then for each $j \in \mathcal{J}$, a fraction π_j of the population holds a quantity a_j of asset a. In this setting, we show in Appendix E that:

Lemma 10. *Starting from a stationary environment, an unexpected freeze of the OTC platform results in a price adjustment on the remaining CEX platform. The sign of the price change is predicted by* $\text{sign}\left(Q^b - \sum_{j\in\mathcal{J}} \pi_j b_j(a_j)\right)$ *before the freeze.*

The intuition behind this result is that if OTC meetings were typically used to reduce positions in liquid assets, then their suspension increases net demand for liquidity on the CEX platform, raising the ETF price.

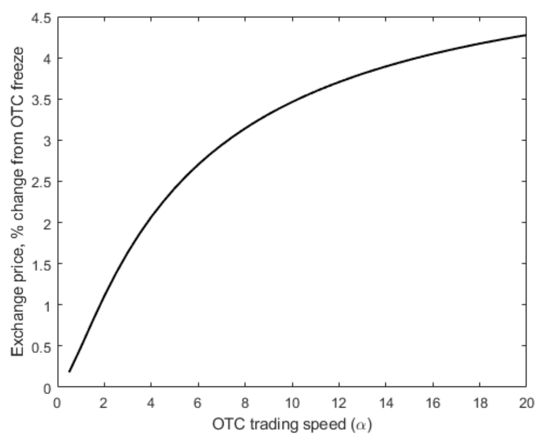


Figure 11: Impact of an OTC freeze on CEX pricing. The x-axis reports the OTC trading speed (α), and the y-axis shows the percentage change in the CEX price following an exogenous halt in OTC trading.

We simulate the model to illustrate the theoretical result. Figure 11 shows an increase in the CEX platform price following an exogenous halt in OTC trading. This effect is consistent with Lemma 10 and reflects the inward shift of liquidity demand. As shown in Figure 6, OTC meetings are typically used to reduce positions in high productivity states, so the suspension of these meetings increases the marginal demand for liquidity on the CEX platform. This price adjustment, derived under the assumption of functioning arbitrage and continuous redemption, illustrates the resilience of ETFs as liquidity vehicles in segmented markets.

However, during the COVID-19 crisis, bond ETFs experienced sharp price declines while trading in their underlying bond markets slowed. The observed drop exceeded the decline in NAV and moved in the opposite direction of the model's prediction. Rather than contradicting the model, this behavior reveals a second pricing mechanism that emerges when arbitrage becomes impaired.

When redemptions are no longer reliably executable, whether due to dealer withdrawal, collateral frictions, or market-wide stress, even a single failed redemption changes the asset-pricing logic. Because ETF shares are fungible, and arbitrage is built on the expectation of parity between market price and NAV, a failed redemption implies that all shares are potentially non-redeemable. In this regime, ETF shares cease to represent liquid claims on a NAV-anchored basket and instead reflect uncertain and potentially impaired terminal value.

This repricing is consistent with rational price discovery. The ETF price no longer tracks the underlying assets, but rather incorporates the value of impaired access and limited liquidity. The resulting discount is not an anomaly but an endogenous response to the structural breakdown of arbitrage. This mechanism aligns with our theoretical results but highlights the fragility of ETF valuation when redemption becomes uncertain.

In sum, while our model predicts upward pricing pressure under a frictionless migration of liquidity, the empirical record suggests that the credibility of the redemption mechanism is pivotal. The ETF discount becomes an endogenous signal of impaired access rather than fundamental deterioration. This insight underscores the dual role of ETFs: as instruments of liquidity and as indicators of intermediation stress.

5.4 Redemption frictions and the liquidity feedback loop in ETFs

During stress periods, ETFs are often perceived as stabilizing instruments: even when the underlying OTC markets become impaired, ETF shares continue to trade. This persistent liquidity is commonly interpreted as evidence that ETFs facilitate price discovery, particularly in illiquid segments of the fixed-income universe. While ETFs may aggregate sentiment when NAVs lag, this surface-level resilience masks a deeper vulnerability: under conditions of market stress, the mechanisms underpinning ETF liquidity begin to unravel.

In our model, this unraveling is driven by redemption frictions that endogenously impair arbitrage. When ETF prices fall below NAV, APs would normally step in to redeem shares and extract the underlying assets. However, because fixed-income ETFs rely on custom redemption baskets rather than pro-rata portfolios, APs anticipate receiving disproportionately illiquid or impaired bonds. When these positions carry large transaction costs, the arbitrage channel weakens. We model this by introducing direction-specific transfer costs, which act as effective balance sheet constraints. These constraints resemble the capital frictions described by Gromb and Vayanos (2018), where the relevant margin becomes the return per unit of constrained intermediary capital. As redemption costs rise, APs exit the arbitrage process and ETF prices diverge further from NAV. Operational details on creation and redemption procedures are summarized in Appendix A.

This breakdown is illustrated in Figure 12, where a negative demand shock interacts with a downward distortion of the redemption frontier. In equilibrium, APs demand a premium to absorb illiquid assets; under stress, this frontier shifts and the discount persists. The wedge between ETF price and NAV becomes a structural feature rather than a temporary anomaly. Consistent with Friewald and Nagler (2019), spreads widen endogenously, and secondary market discounts reflect not mispricing, but impaired intermediation capacity. In parallel, stale NAVs in bond portfolios can delay valuations and generate dilution or run incentives, further weakening the NAV anchor during stress (Choi et al., 2022).

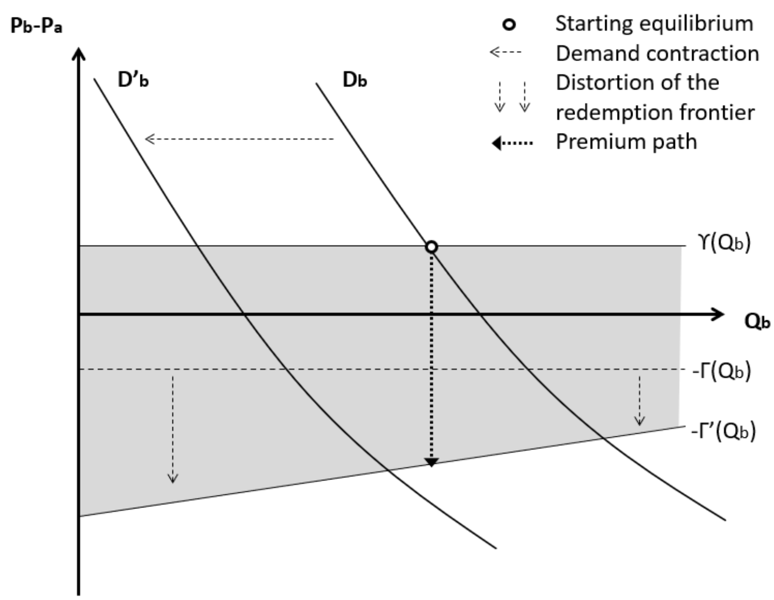


Figure 12: Impact of a negative demand shock with redemption-frontier distortion. The contraction in CEX demand is denoted by D'_b, and the redemption frontier is schematically distorted from $\Gamma(Q_b)$ to $\Gamma'(Q_b)$. Starting from the original equilibrium on the expansion frontier $\Upsilon(Q_b)$, the diagram illustrates the resulting downward adjustment in the ETF premium.

Beyond this modeled mechanism, the distorted redemption frontier depicted in Figure 12 also reflects a broader fragility that can arise when redemption flows overwhelm the capacity of OTC markets. Under normal conditions, OTC and CEX platforms operate under relatively stable volumes. During stress, however, redemptions can abruptly force concentrated outflows of illiquid assets onto OTC venues, shifting supply conditions in ways that exceed dealer absorption capacity. This unanticipated adjustment compounds the initial demand contraction and can induce fire-sale dynamics that further depress prices and destabilize liquidity conditions. While not formally modeled, this interaction between redemption flows and OTC imbalance constitutes an additional amplification channel consistent with the asymmetries embedded in the redemption process.

Our model captures the endogenous liquidity spiral that arises when arbitrage breaks down and redemptions persist. These dynamics are consistent with behavior observed during the March 2020 market dislocation (Figure 3), when bond ETFs traded at sustained discounts of 5–6%. ETF prices led NAVs downward, in line with evidence that ETFs incorporate bad news faster than evaluated bond portfolios (O'Hara and Zhou, 2021). The deterioration was exacerbated by redemption basket design: to minimize their own transaction costs, ETF sponsors concentrated illiquid or hard-to-value assets in the baskets they handed off to APs, as shown empirically in Koont et al. (2025). Arbitrage incentives eroded further, and persistent dislocations emerged.

While central bank interventions eventually restored market functioning, these pricing gaps should not be seen as transient arbitrage anomalies. As Huang et al. (2025) emphasize, intermediation frictions, rather than credit risk, explain much of the observed widening in credit spreads during stress. In this view, ETF discounts are not failures of ETF design, but equilibrium outcomes in a world where secondary market liquidity is endogenous, path dependent, and structurally fragile.

The implications extend beyond ETF-specific mechanisms. As Greenwood and Thesmar (2011) argue, overlapping portfolios and common ownership amplify fire-sale contagion when liquidity shocks propagate across vehicles. In our setting, this is reflected in the path dependence of redemption dynamics. Once APs withdraw and redemptions distort basket composition, the system can transition into a persistent state of dislocation, in which price distortions remain even after the initial shock has passed. The result is a slow-moving crisis: prices reflect sentiment and structural segmentation more than fundamental value.

Practitioner commentary reinforces this view. Firms such as Invesco and VanEck caution against interpreting ETF discounts in crisis as arbitrage opportunities. Instead, they emphasize liquidity opacity, structural valuation gaps, and redemption mechanics as the primary drivers of discount behavior.[31] Recent volatility in ETFs exposed to structured credit, including CLOs, highlights how the wrapper design can both enable liquidity and transmit instability.

In sum, redemption frictions generate a nonlinear feedback loop that drives structural discounts in equilibrium and risks overwhelming OTC markets out of equilibrium. The model and evidence point to a central insight: ETF-based liquidity transformation is inherently fragile when backed by illiquid assets, calling for a reassessment of ETF structure and systemic safeguards.

6 Discussion and Conclusion

This article develops a continuous-time framework to analyze how fixed-income ETFs interact with OTC market structures to shape asset pricing, liquidity dynamics, and systemic fragility. Motivated by the growing footprint of ETFs that transform illiquid exposures into liquid claims, particularly in credit markets, the model formalizes trade segmentation across two frictionally connected venues. It connects theoretical mechanisms of liquidity provision and intermediation to observed patterns of ETF price behavior during stress episodes such as the COVID-19 crisis.

The contributions of this paper are threefold. First, we construct a dual-platform model in which investors allocate across an illiquid OTC asset and a continuously tradable ETF substitute, while intermediaries transfer assets between platforms at asymmetric, state-dependent costs. The framework nests equilibrium price formation, investor heterogeneity, and dynamic arbitrage mechanics in a tractable structure that yields closed-form solutions. Second, we show that ETF growth introduces a trade-off: while it enhances access and liquidity, it may alter the economics of OTC intermediation in ways that reduce the benefits to sustain decentralized market-making. Our simulations reveal that this reallocation can result in path-dependent equilibria, shifting the long-run structure of market depth and intermediation. Third, we identify a feedback loop between ETF redemption mechanics and OTC frictions. Under stress, even moderate mismatches between ETF liquidity and asset backing can trigger a dislocation spiral, replicating stylized facts from the COVID-19 ETF discounts.

From a theoretical perspective, our model contributes to the literature by endogenizing ETF market size and platform segmentation in response to investor preference shocks and platform-specific frictions. The illiquidity hedging mechanism embedded in the exchange op-

[31] See Invesco (2024), "Understanding ETF trading and liquidity: Arbitrage, premiums, and discounts," available at https://www.invesco.com/apac/en/institutional/insights/etf/understanding-etf-trading-and-liquidity-etf-arbitrage-premiums-and-discounts.html.

tion resembles the role of spare liquidity in dynamic portfolio choice models with intertemporal optimization. Yet unlike standard consumption-smoothing models, liquidity in our setting is not simply a store of value, it is an instrument of mobility across platforms. This distinction is key to understanding how segmented markets evolve when liquidity shocks hit only part of the system.

At the same time, our findings underscore the virtues of ETFs in transforming access to credit markets. By enabling continuous tradability of exposures that would otherwise be siloed in bilateral trades, ETFs expand the investable universe, facilitate risk transfer, and support price discovery, particularly when OTC markets are impaired. This role was apparent during the COVID-19 crisis, where ETFs remained active even as bond markets froze, helping to support price discovery. Recent evidence shows that liquidity seekers systematically concentrate in ETFs, enabling fee premia and venue-specific price setting, which underscores the welfare gains from improved tradability (Khomyn et al., 2024). Our model formalizes this stabilizing channel: ETFs absorb demand spillovers under moderate stress and allow investors to partially hedge preference shocks through platform substitution. In this sense, ETFs are not merely wrappers; they are liquidity engines.

However, the packaging of illiquid credit assets into ETFs embeds structural fragility by decoupling liquidity from the valuation of the underlying assets. Custom redemption baskets, endogenous arbitrage incentives, and imbalances in platform liquidity provision can give rise to persistent pricing dislocations. Our model simulations demonstrate that even in the absence of fundamental deterioration, ETF redemptions may accelerate, price premia may widen, and intermediary incentives may weaken, all amplifying stress through internal dynamics. This feature aligns with evidence that arbitrage breaks down when balance sheet constraints tighten or valuations become opaque (O'Hara and Zhou, 2021). In our model, these limits are not exogenous, they emerge from platform interactions.

From a policy perspective, our findings support a more nuanced view of ETF resilience. We show that ETF platforms can stabilize markets through accessibility and flexibility, but they also introduce a layer of system dynamics that may amplify stress through feedback loops. The systemic risk is not in ETF popularity per se, it lies in the asymmetric burden placed on redemption mechanics and in the weakening of the slower, shock-absorbing OTC layer. In this light, regulation should not focus solely on ETF leverage or NAV transparency, but also on redemption design and intermediation capacity. As emphasized by Faria-E-Castro et al. (2017), transparency is not a sufficient safeguard; without credible backstops it can exacerbate fragility. The parallel suggests that ETF regulation must balance disclosure with robust support mechanisms, ensuring that the benefits of transparency are not offset by destabilizing dynamics. Mandating pro-rata redemption structures, incentivizing standardized basket disclosure, or aligning AP incentives across stress states may reduce the likelihood of fragility transmission. These measures are consistent with suggestions in Grill et al. (2018) and could be tested in stress simulation regimes.

Several paths remain open for future research. First, future work could investigate optimal platform regulation under endogenous AP participation or withdrawal. A second extension would allow for nonstationary preference shocks or information asymmetries, linking our framework to the broader literature on adverse selection and dynamic market unraveling (Chiu and Koeppl, 2016). Lastly, future research should examine how ETF dislocations interact with other investment vehicles holding correlated credit exposures. Our model highlights how liquidity transformation affects price behavior under stress, but in practice, redemptions may become synchronized across ETFs and mutual funds, amplifying volatility through joint asset sales and portfolio overlap. Measuring the systemic consequences of these cross-vehicle spillovers remains a critical and policy-relevant task.

ETFs have become central to the functioning of modern financial markets. They democratize access, streamline execution, and improve price visibility in complex asset classes. But when the appearance of liquidity masks deeper segmentation, the very features that make ETFs valuable can also become the channels through which stress propagates. Understanding this duality is not only essential to interpreting market behavior in crisis, but also to designing resilient financial infrastructure.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). "Measuring Systemic Risk." *Review of Financial Studies* 30.1, 2–47.

Amihud, Y. and Mendelson, H. (1980). "Dealership market: Market-making with inventory." *Journal of Financial Economics* 8.1, 31–53.

Ben-David, I., Franzoni, F., Kim, B., and Moussawi, R. (2023). "Competition for Attention in the ETF Space." *The Review of Financial Studies* 36.3, 987–1042.

Ben-David, I., Franzoni, F., and Moussawi, R. (2018). "Do ETFs Increase Volatility?" *The Journal of Finance* 73.6, 2471–2535.

Brown, D. C., Davies, S. W., and Ringgenberg, M. C. (2021). "ETF Arbitrage, Non-Fundamental Demand, and Return Predictability." *Review of Finance* 25.4, 937–972.

Brunnermeier, M. K. and Pedersen, L. H. (2009). "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22.6, 2201–2238.

Carapella, F. and Monnet, C. (2020). "Dealers' insurance, market structure, and liquidity." *Journal of Financial Economics* 138.3, 725–753.

Chiu, J. and Koeppl, T. V. (2016). "Trading Dynamics with Adverse Selection and Search: Market Freeze, Intervention and Recovery." *The Review of Economic Studies* 83.3, 969–1000.

Choi, J., Kronlund, M., and Oh, J. Y. J. (2022). "Sitting bucks: Stale pricing in fixed income funds." *Journal of Financial Economics* 145.2, 296–317.

Coval, J., Jurek, J., and Stafford, E. (2009). "The Economics of Structured Finance." *Journal of Economic Perspectives* 23.1, 3–25.

Dannhauser, C. D. and Hoseinzade, S. (2021). "The Unintended Consequences of Corporate Bond ETFs: Evidence from the Taper Tantrum." *The Review of Financial Studies* 35.1, 51–90.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2005). "Over-the-Counter Markets." *Econometrica* 73.6, 1815–1847.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2007). "Valuation in Over-the-Counter Markets." *The Review of Financial Studies* 20.6, 1865–1900.

Duffie, D. and Sun, Y. (2007). "Existence of Independent Random Matching." *The Annals of Applied Probability* 17.1, 386–419.

Duffie, D. and Zhu, H. (2011). "Does a Central Clearing Counterparty Reduce Counterparty Risk?" *Review of Asset Pricing Studies* 1.1, 74–95.

Dugast, J., Üslü, S., and Weill, P.-O. (2022). "A Theory of Participation in OTC and Centralized Markets." *The Review of Economic Studies* 89.6, 3223–3266.

ECB (2018). "Financial Stability Review, November 2018." *Financial Stability Review*.

Evans, M. D. D. and Lyons, R. K. (2002). "Order Flow and Exchange Rate Dynamics." *Journal of Political Economy* 110.1, 170–180.

Faria-E-Castro, M., Martinez, J., and Philippon, T. (2017). "Runs versus Lemons: Information Disclosure and Fiscal Capacity." *The Review of Economic Studies* 84.4, 1683–1707.

Financial Conduct Authority, Mittendorf, D., Neumeier, C., O'Neill, P., and Rahimi, K. (2021). *Research Note: Capital market liquidity in the 2020 coronavirus crisis*. Financial Conduct Authority.

Friewald, N. and Nagler, F. (2019). "Over-the-Counter Market Frictions and Yield Spread Changes." *The Journal of Finance* 74.6, 3217–3257.

Gârleanu, N. (2009). "Portfolio choice and pricing in illiquid markets." *Journal of Economic Theory* 144.2, 532–564.

Gârleanu, N. and Pedersen, L. H. (2016). "Dynamic portfolio choice with frictions." *Journal of Economic Theory* 165, 487–516.

Glode, V. and Opp, C. C. (2020). "Over-the-Counter versus Limit-Order Markets: The Role of Traders' Expertise." *The Review of Financial Studies* 33.2, 866–915.

Glode, V., Opp, C. C., and Sverchkov, R. (2022). "To pool or not to pool? Security design in OTC markets." *Journal of Financial Economics* 145.2, 508–526.

Glosten, L., Nallareddy, S., and Zou, Y. (2021). "ETF Activity and Informational Efficiency of Underlying Securities." *Management Science* 67.1, 22–47.

Goldstein, I., Jiang, H., and Ng, D. T. (2017). "Investor flows and fragility in corporate bond funds." *Journal of Financial Economics* 126.3, 592–613.

Gorton, G. and Metrick, A. (2012). "Securitized banking and the run on repo." *Journal of Financial Economics* 104.3, 425–451.

Greenwood, R. and Thesmar, D. (2011). "Stock price fragility." *Journal of Financial Economics* 102.3, 471–490.

Grill, M., Lambert, C., Marquardt, P., Watfe, G., and Weistroffer, C. (2018). "Counterparty and liquidity risks in exchange-traded funds." *ECB Financial Stability Review*.

Gromb, D. and Vayanos, D. (2018). "The Dynamics of Financially Constrained Arbitrage." *The Journal of Finance* 73.4, 1713–1750.

Grossman, S. J. and Miller, M. H. (1988). "Liquidity and Market Structure." *The Journal of Finance* 43.3, 617–633.

Guerrieri, V. and Shimer, R. (2014). "Dynamic Adverse Selection: A Theory of Illiquidity, Fire Sales, and Flight to Quality." *American Economic Review* 104.7, 1875–1908.

Hauser, A. (2021). “From Lender of Last Resort to Market Maker of Last Resort via the dash for cash: why central banks need new tools for dealing with market dysfunction.” *Bank of England Central bank speech*.

He, Z. and Song, Z. (2025). “Agency MBS as Safe Assets.” *The Review of Financial Studies*.

Hendershott, T., Li, D., Livdan, D., and Schürhoff, N. (2024). “When failure is an option: Fragile liquidity in over-the-counter markets.” *Journal of Financial Economics* 157, 103859.

Hendershott, T., Menkveld, A. J., Praz, R., and Seasholes, M. (2022). “Asset Price Dynamics with Limited Attention.” *The Review of Financial Studies* 35.2, 962–1008.

Holden, C. W. and Nam, J. (2024). “Market accessibility, bond ETFs, and liquidity.” *Review of Finance* 28.5, 1725–1758.

Holden, C. W., Lu, D., Lugovskyy, V., and Puzzello, D. (2021). “What is the impact of introducing a parallel OTC market? Theory and evidence from the chinese interbank FX market.” *Journal of Financial Economics* 140.1, 270–291.

Huang, J.-Z., Nozawa, Y., and Shi, Z. (2025). “The Global Credit Spread Puzzle.” *The Journal of Finance* 80.1, 101–162.

Kargar, M., Lester, B., Lindsay, D., Liu, S., Weill, P.-O., and Zúñiga, D. (2021). “Corporate Bond Liquidity during the COVID-19 Crisis.” *The Review of Financial Studies* 34.11, 5352–5401.

Khomyn, M., Putniņsˇ, T., and Zoican, M. (2024). “The Value of ETF Liquidity.” *The Review of Financial Studies* 37.10, 3092–3148.

Koont, N., Ma, Y., Pastor, L., and Zeng, Y. (2025). “Steering a Ship in Illiquid Waters: Active Management of Passive Funds.” *The Review of Financial Studies*.

Kyle, A. S. (1985). “Continuous Auctions and Insider Trading.” *Econometrica* 53.6, 1315–1335.

Lagos, R. and Rocheteau, G. (2007). “Search in Asset Markets: Market Structure, Liquidity, and Welfare.” *American Economic Review* 97.2, 198–202.

Lagos, R. and Rocheteau, G. (2009). “Liquidity in Asset Markets With Search Frictions.” *Econometrica* 77.2, 403–426.

Lester, B., Rocheteau, G., and Weill, P.-O. (2015). “Competing for Order Flow in OTC Markets.” *Journal of Money, Credit and Banking* 47 (S2), 77–126.

Lettau, M. and Madhavan, A. (2018). “Exchange-Traded Funds 101 for Economists.” *Journal of Economic Perspectives* 32.1, 135–154.

Liu, X. (2016). “Interbank Market Freezes and Creditor Runs.” *The Review of Financial Studies* 29.7, 1860–1910.

Malamud, S. and Rostek, M. (2017). “Decentralized Exchange.” *The American Economic Review* 107.11, 3320–3362.

Marta, T. and Riva, F. (2025). “Do ETFs increase the comovements of their underlying assets? Evidence from a switch in ETF replication technique.” *Journal of Banking & Finance* 170, 107333.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press. 998 pp.

Melin, L. (2012). “Multi-Speed Markets.” *Mimeo*.

Municipal Securities Rulemaking Board (2022). *Trends in Municipal Securities Ownership*. Washington, DC: Municipal Securities Rulemaking Board.

O’Hara, M. and Zhou, X. (2021). “Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis.” *Journal of Financial Economics* 142.1, 46–68.

Pagnotta, E. S. and Philippon, T. (2018). “Competing on Speed.” *Econometrica* 86.3, 1067–1115.

Pinter, G., Wang, C., and Zou, J. (2024). “Size Discount and Size Penalty: Trading Costs in Bond Markets.” *The Review of Financial Studies* 37.7, 2156–2190.

Rahi, R. and Zigrand, J.-P. (2009). “Strategic Financial Innovation in Segmented Markets.” *Review of Financial Studies* 22.8, 2941–2971.

Rust, J. and Hall, G. (2003). “Middlemen versus Market Makers: A Theory of Competitive Exchange.” *Journal of Political Economy* 111.2, 353–403.

Schwert, M. (2017). “Municipal Bond Liquidity and Default Risk.” *The Journal of Finance* 72.4, 1683–1722.

Shleifer, A. and Vishny, R. W. (1997). “The Limits of Arbitrage.” *The Journal of Finance* 52.1, 35–55.

U.S. Securities and Exchange Commission (2019). *Exchange-Traded Funds: Small Entity Compliance Guide (Rule 6c-11)*. U.S. Securities and Exchange Commission, Division of Investment Management.

Üslü, S. (2019). “Pricing and Liquidity in Decentralized Asset Markets.” *Econometrica* 87.6, 2079–2140.

Vayanos, D. and Vila, J.-L. (2021). “A Preferred-Habitat Model of the Term Structure of Interest Rates.” *Econometrica* 89.1, 77–112.

Appendix

A ETF Creation and Redemption Mechanics

This appendix outlines the operational structure of ETF share creation and redemption in fixed-income markets, emphasizing the role of custom baskets and the incentives of APs. Understanding these mechanics is crucial for interpreting the model's fragility results. We avoid duplicating basic ETF descriptions covered in Section 2, instead focusing on the specific institutional features that distinguish bond-backed ETFs.

Primary vs. secondary markets

An ETF operates through two linked markets. In the secondary market, investors trade ETF shares on an exchange (centralized market) just like stocks. In the primary market, specialized intermediaries known as APs interact directly with the ETF sponsor (issuer) to create or redeem shares in large blocks (often called "creation units"). Creation involves an AP delivering a specified basket of assets (and/or cash) to the fund in exchange for newly issued ETF shares, which the AP can then resell to investors. Redemption is the reverse: the AP returns a block of ETF shares to the issuer and receives in exchange a basket of the fund's underlying assets. Through this arbitrage process, APs help align the ETF's market price with the value of its portfolio: if the ETF price is above NAV, APs can profit by creating shares and selling them at the premium, and if the ETF price is below NAV, they can redeem shares and extract underlying bonds of equal NAV. These primary-market trades are typically executed in-kind rather than in cash, especially for bond ETFs, to avoid forcing the fund to transact in illiquid markets. Under normal conditions, the threat of AP arbitrage keeps ETF prices closely tethered to NAV.

Custom Baskets and U.S. Securities and Exchange Commission (SEC) Rule 6c-11

Unlike traditional index mutual funds, ETFs are not required to transfer a pro-rata slice of their entire portfolio in creations or redemptions. Rule 6c-11 (adopted by the SEC in September 2019)[32] formally permits ETF issuers to employ *custom baskets* that can deviate from the fund's index composition, provided they follow transparent policies. This flexibility is particularly important for fixed-income ETFs, where fully replicating an index is impractical and many bonds trade infrequently. In practice, ETF issuers construct a creation and redemption basket each day and communicate their contents to APs prior to market open. These baskets

[32] See U.S. Securities and Exchange Commission (2019).

typically contain a subset of the fund's holdings and sometimes a cash component, designed such that the total value matches the NAV of a creation unit.

Custom baskets are central to ETF liquidity management: rather than include hundreds of bonds in index weights, bond ETF baskets usually contain only the more liquid portion of the portfolio and use cash or equivalents to make up value. Empirical evidence shows that an average corporate bond ETF basket might contain only 20–30% of holdings by name, skewed toward large, tradable bonds, plus a cash balance. Koont et al. (2025) document that, on average, U.S. investment-grade bond ETF baskets contain 5–12% cash and only a small fraction of index constituents. This represents a trade-off: including cash and limiting baskets to liquid bonds reduces AP transaction costs, but increases tracking error. In short, bond ETFs actively manage baskets to balance liquidity provision and index tracking, and as the authors argue, such management is necessary to transform liquidity while still roughly replicating portfolios.

AP–issuer basket negotiation

The content of creation and redemption baskets is not set unilaterally but results from mutual incentives. APs prefer baskets composed of assets that are easy to source or dispose of. When creating shares, APs gravitate toward delivering liquid bonds or cash. Issuers, in contrast, may use redemptions to offload illiquid or overweight bonds. As a result, creation baskets tend to favor liquid instruments, while redemption baskets may include illiquid or riskier positions. Industry practice reflects this asymmetry: APs often have latitude in creation baskets, whereas redemption baskets are closer to a one-way offer from the fund. This helps ETFs shed unwanted exposures during outflows but can strain AP engagement under stress. If volatility spikes and redemption baskets contain hard-to-sell bonds, APs demand higher compensation (via larger ETF discounts) or exit arbitrage altogether. This misalignment was evident in March 2020, when many APs scaled back activity, fearing redemption baskets would contain the most impaired bonds.[33]

Effects on price formation and NAV dynamics

The creation and redemption mechanism normally keeps ETF market prices close to portfolio value. However, in fixed-income ETFs several features can cause persistent price dislocations:

Stale bond pricing vs. real-time ETF prices. Many bond ETFs hold securities that do not trade daily. NAVs are computed using pricing services or models, which can lag true

[33] See Financial Conduct Authority (FCA), May 2021 market commentary, available at: Financial Conduct Authority (FCA): Capital market liquidity in the 2020 coronavirus crisis.

values. ETF shares, by contrast, trade continuously. During the COVID-19 selloff, ETF prices dropped sharply to anticipated fire-sale levels, while NAVs adjusted slowly.[34] Studies confirm ETF prices were more reactive than NAVs, reflecting price discovery rather than arbitrage slack.

Transaction costs and arbitrage bands. Arbitrage is not frictionless. APs incur costs (bid–ask spreads, impact, financing), and will only act if discrepancies exceed these costs. In stress, costs rise sharply, widening no-arbitrage bands. Discounts around 5–6% in March 2020 persisted because arbitrage was uneconomic, given the risk of receiving illiquid assets in redemption baskets, as emphasized in the FCA note and contemporaneous policy commentary.

NAV tracking and portfolio composition. Custom baskets and cash can create divergence between fund portfolios and their indices. If creations occur with cash, the fund may temporarily hold cash rather than bonds. If redemptions remove subsets of bonds, the fund's composition drifts. Koont et al. (2025) show bond ETFs in less-liquid markets tolerate larger tracking error by holding more cash, an optimal response to high transaction costs. In turbulence, these adjustments slow or cease, so ETF NAVs may not fully reflect the hypothetical mark-to-market of the index. Prices incorporate discounts for illiquidity, while NAVs can lag behind.

Interpretation and implications

ETF creations and redemptions enable liquidity transformation: investors obtain a continuously tradable instrument even when the underlying bonds are illiquid. Custom baskets are central to this transformation, allowing transactions in manageable subsets of assets and use of cash buffers. While this improves feasibility, the same design creates fragility. In stress, stale NAVs, high costs, and adverse selection can lead to persistent ETF discounts. Such discounts often reflect structural stress, not arbitrage opportunities. APs may withdraw temporarily, and the ETF's liquidity advantage becomes a transmission channel for distress. The interactions of these forces are analyzed in Section 5.4 of the main text.

B Data Construction and Empirical Foundations

B.1 Data sources

We obtain fund-level data from the CRSP Survivor-Bias-Free U.S. Mutual Fund Database, using two primary datasets: (i) the *Daily Returns and Net Asset Values file* (`daily_nav_ret`),

[34] See again FCA (2021).

and (ii) the *Fund Summary – Quarterly file* (fund_summary_q). The daily NAV/returns file provides each fund's daily NAV per share (dnav) and daily total return (dret), along with identifiers such as ticker, CUSIP, and shares outstanding. The fund summary file contains classification and composition details, including CRSP objective codes (crsp_obj_cd), Lipper classifications (lipper_class_name, lipper_asset_cd), and quarterly percentage portfolio allocations (e.g., per_com, per_corp, per_muni, per_govt, per_abs, per_mbs). These variables are used to identify fund categories and to filter the ETF universe.

The database flags fund structures via the ETF/exchange-traded notes (ETNs) indicator (et_flag), which distinguishes ETFs (et_flag = ''F'') from ETNs or mutual funds. We restrict our analysis to ETFs only. All data were extracted as of December 2024. The sample covers daily observations from January 1, 2020 through December 31, 2024. Interest rate benchmarks, specifically the 90-day SOFR, are obtained from the Federal Reserve Bank of St. Louis FRED database (series: SOFR90DAYAVG).

B.2 Sample construction and filtering

Starting from the full CRSP universe, we restrict to ETFs using the et_flag = ''F'' indicator. We then remove equity ETFs and hybrid allocation funds to focus exclusively on fixed-income ETFs. Specifically, funds classified with lipper_asset_cd = ''EQ'' are excluded, as are funds with portfolio compositions showing more than 50 percent invested in common or preferred stock (per_com, per_pref). We also remove funds with crsp_obj_cd beginning with "OC," which correspond to currency ETFs, cryptocurrency products (e.g., Bitcoin and Ethereum ETFs), and certain non-fixed income multi-asset strategies. After excluding all EQ funds and OC-coded funds, we apply a single fund-name screen for "Large Cap", which removes exactly one remaining observation.

To ensure that no residual equity-style ETFs remain, we also exclude the single observation with crsp_obj_cd = ''EDYI''. This code denotes an equity, domestic, style-classified fund and is therefore outside our fixed-income universe.

From the remaining sample, we classify ETFs into five mutually exclusive fixed-income categories:

1. **Government Bond ETFs:** Funds investing at least 90% of assets in government bonds (per_govt), excluding Treasury Inflation-Protected Securities (TIPS)-dedicated funds.
2. **Corporate Bond ETFs:** Funds with primary exposure to corporate bonds, identified by high allocations to corporate debt and objective codes for investment-grade or high-yield credit.

3. **Municipal Bond ETFs:** Funds identified via Lipper's municipal asset code (`MB`) or objective codes for tax-exempt fixed income.

4. **ABS ETFs:** Funds identified by the CRSP objective code `OM`. Within CRSP's classification system, `O` denotes "Other," and the subclass `M` specifies "Mortgage-backed." Thus, `OM` corresponds to "Other, Mortgage-backed" funds. This category encompasses ETFs whose primary holdings are securitized and structured fixed-income instruments, including MBS, commercial MBS, ABS, senior secured loans, floating-rate credit instruments, and related products. All CLO ETFs in the CRSP database are classified within this `OM` group.

5. **Other Non-Government Bond ETFs:** A residual category for fixed-income ETFs that do not meet the above criteria, including international bond funds, TIPS-dedicated ETFs, and diversified multi-sector funds.

6. **CLO ETFs (subset of ABS):** For the purpose of the Appendix, we separately identify a tractable subset of ABS ETFs that focus on CLOs. These funds are drawn from the CRSP objective code `OM` and confirmed by fund names and cross-checking. Because the number of CLO ETFs is still limited, this separation is feasible. The CRSP sample features exactly 10 CLO ETFs, with tickers: AAA, CLOA, CLOI, CLOX, CLOZ, HSRT, ICLO, JAAA, JBBB, PAAA. Further details on market growth and liquidity trends are provided in Appendix C.

Funds are classified into five mutually exclusive fixed-income categories (1–5) for the main body of the paper. For the Appendix, we additionally feature CLO ETFs as a tractable subset of the ABS category. This ensures that the main classification used in Section 2 remains consistent, while also allowing us to highlight the small set of CLO ETFs separately for descriptive purposes. We de-duplicate by CUSIP and ticker to retain a single record per ETF before merging prices and NAVs. The ETF-level classification file is merged by CUSIP to the CRSP daily file containing prices, shares outstanding, and NAVs.

B.3 Variable definitions and aggregation

Let P_{it} denote the reported closing price for ETF i on date t (`abs(PRC)`), taken in absolute value. Let NAV_{it} be the NAV per share (`dnav`), and let $Shares_{it}$ be shares outstanding (`SHROUT`, recorded in thousands and multiplied by 1,000). We define ETF-level market and NAV capitalizations as

$$\text{MarketCap}_{it} = P_{it} \times Shares_{it}, \qquad \text{NAVCap}_{it} = NAV_{it} \times Shares_{it}.$$

The ETF-level premium or discount in percent is

$$\text{PremDisc}_{it} = \frac{\text{MarketCap}_{it} - \text{NAVCap}_{it}}{\text{NAVCap}_{it}} \times 100.$$

This construction measures how far the ETF's total market value deviates from the valuation implied by its NAV for the same number of shares. Using capitalization rather than per-share ratios is equivalent at the fund level, but it is crucial for aggregation: at the category-by-date level we first *sum* market and NAV capitalizations across all ETFs in the category, then compute

$$\text{PremDisc}_{ct} = \frac{\sum_{i\in c} \text{MarketCap}_{it} - \sum_{i\in c} \text{NAVCap}_{it}}{\sum_{i\in c} \text{NAVCap}_{it}} \times 100,$$

which yields a market-cap–weighted premium/discount for category c on date t. This avoids averaging fund-level premia equally and ensures that larger ETFs receive proportionally greater weight. A positive value indicates that the ETF is trading at a premium relative to its NAV, while a negative value indicates that it is trading at a discount relative to NAV. These measures are computed daily from January 2020 through December 2024. For analysis, data are aggregated to the ETF category-by-date level. Category-level market capitalization is obtained by summing across all ETFs in the group. Category-level premium/discount is computed as the average across constituent ETFs.

C CLO ETFs: market growth and liquidity trends

This section provides descriptive evidence on the rise of CLO ETFs. While the core model abstracts from specific asset classes, these instruments offer a relevant real-world illustration of the liquidity transformation mechanism analyzed in the paper. CLO ETFs provide secondary-market access to structured credit positions that are otherwise traded OTC. The figure below documents the evolution of aggregate market capitalization and trading volume for U.S.-listed CLO ETFs between 2020 and 2024. Both measures remained minimal through 2022: total market capitalization stayed below $1 billion, and trading activity was close to zero. Beginning in 2023, both series accelerated markedly. By the end of 2024, aggregate market capitalization approached $23 billion, and monthly trading volume exceeded $20 billion. Over the same period, short-term interest rates, measured here by the 90-day SOFR, rose significantly and then stabilized. While the figure does not establish causality, it highlights the concurrent growth in ETF-based access to traditionally illiquid credit markets.

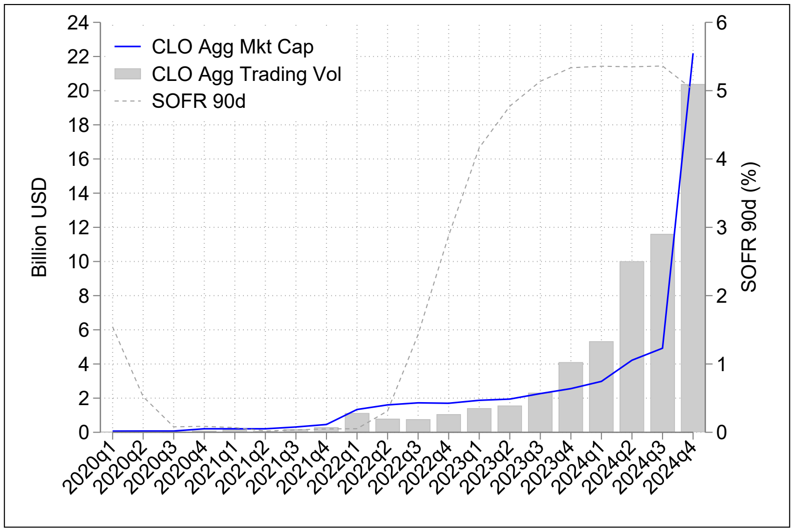


Figure A1: Market Capitalization, Trading Volume, and 90-Day SOFR for CLO ETFs, 2020–2024. This figure presents quarterly data for U.S.-listed CLO ETFs, including aggregate market capitalization (blue line, right axis, in billion USD), total trading volume (gray bars, left axis, in billion USD), and the 90-day SOFR (gray dashed line, right axis, in percent). CLO ETF data are based on aggregated fund-level statistics. The sample spans the period from January 2020 through December 2024.

D Model derivations

Bellman equation

The value function at time t for an agent facing preference state $i(t) = i$ and holding portfolio $(a, b_i(a,t))$[35] is:

$$V_i(a,t) = \mathbb{E}_t \left[\begin{array}{c} \int_t^{\widetilde{\tau}_m} e^{-r(s-t)} \left[u_{i(s)}(a, b_{i(s)}(a,s)) - p_b(s) \left[b_{i(s)}(a,s) - b_{i(s^-)}(a,s^-) \right] \right] ds \\ +e^{-r(\widetilde{\tau}_m - t)} \left\{ \begin{array}{c} V_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m), \widetilde{\tau}_m) \\ -\left[p_a(\widetilde{\tau}_m) + \varphi_{i(\widetilde{\tau}_m)}(a, \widetilde{\tau}_m) \right] \left[a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m) - a \right] \\ -p_b(\widetilde{\tau}_m) \left[b_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m), \widetilde{\tau}_m) - b_{i(\widetilde{\tau}_m)}(a, \widetilde{\tau}_m) \right] \end{array} \right\} \end{array} \right]$$

$$\begin{aligned} V_i(a,t) = \mathbb{E}_t \Bigg[\int_t^{\widetilde{\tau}_m} & e^{-r(s-t)} \left[u_{i(s)}(a, b_{i(s)}(a,s)) - p_b(s) \left[b_{i(s)}(a,s) - b_{i(s^-)}(a,s^-) \right] \right] ds \\ & + e^{-r(\widetilde{\tau}_m - t)} \Big\{ V_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m), \widetilde{\tau}_m) - \left[p_a(\widetilde{\tau}_m) + \varphi_{i(\widetilde{\tau}_m)}(a, \widetilde{\tau}_m) \right] \left[a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m) - a \right] \\ & \qquad - p_b(\widetilde{\tau}_m) \left[b_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m), \widetilde{\tau}_m) - b_{i(\widetilde{\tau}_m)}(a, \widetilde{\tau}_m) \right] \Big\} \Bigg] \end{aligned}$$

with $\widetilde{\tau}_m$ the stopping time corresponding to a meeting Poisson arrival.

In words, the first component is composed of the utility flow from the current point in time

[35] $b_i(a,t)$ is the optimal liquid holding given the illiquid position and the current type.

until the next OTC meeting. It encompasses the possibility for the liquid asset to get traded during that time. The second term is the discounted continuation value when a meeting arises. This term comprises the new after-meeting value, net from the fees paid to access the OTC trading platform[36], and the costs to readjust both portfolio components.

After integrating by part[37] the first term, the value function $V_i(a,t)$ becomes

$$\mathbb{E}_t\left[\begin{array}{c} p_b(t)b_i(a,t)-e^{-r(\widetilde{\tau}_m-t)}p_b(\widetilde{\tau}_m)b_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m) \\ +\int_t^{\widetilde{\tau}_m} e^{-r(s-t)}\left[u_{i(s)}(a,b_{i(s)}(a,s))-[rp_b(s)-\dot{p}_b(s)]\,b_{i(s)}(a,s)\right]ds \\ +e^{-r(\widetilde{\tau}_m-t)}\left\{\begin{array}{c} V_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m),\widetilde{\tau}_m)-\left[p_a(\widetilde{\tau}_m)+\varphi_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m)\right]\left[a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m)-a\right] \\ -p_b(\widetilde{\tau}_m)\left[b_{i(\widetilde{\tau}_m)}(a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m),\widetilde{\tau}_m)-b_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m)\right]\end{array}\right\}\end{array}\right]$$

Exchange trades

At each and every point in time, for $i(t)=i$ the optimal response function $b_i(a,t)$ is defined implicitly from the maximization

$$\begin{aligned} b_i(a,t) &= \arg\max_{b'\geq 0}\left[u_i(a,b')-q_b(t)b'\right] \\ \text{with } q_b(t) &= rp_b(t)-\dot{p}_b(t)\end{aligned}$$

leading to the first-order condition:

$$\partial_2 u_i(a,b_i(a,t))\leq q_b(t),\ \text{with equality for } b_i(a,t)>0$$

The optimal holding of liquid asset is defined at any point in time, conditional on the holding of the illiquid asset and the realization of the preference shock. We then define the trade gain (or net utility flow)

$$\widehat{u}_i(a,t)\equiv u_i(a,b_i(a,t))-q_b(t)b_i(a,t)$$

which captures the utility flow captured by an investor who has access to an exchange platform and takes advantage of it to keep its holding of the liquid asset at the optimal level.

Three state variables emerge at this point: a the level of illiquid asset, i the preference shock and t the time dependency. Time dependency is not redundant vis-a-vis the preference shock because of the role for the liquid asset holding cost $q_b(t)$.

[36]The per-unit fees $\varphi_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m)$ are mapped to the total transaction fees $\phi_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m)=\varphi_{i(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m)\left[a_{i(\widetilde{\tau}_m)}(\widetilde{\tau}_m)-a\right]$.

[37]Details of the integration by part are

$$\begin{aligned}\int_t^{\widetilde{\tau}_m} e^{-r(s-t)}\left[u_{k(s)}(a,b_{k(s)}(a,s))-p_b(s)\dot{b}_{k(s)}(a,s)\right]ds \\ = p_b(t)b_{k(t)}(a,t)-e^{-r(\widetilde{\tau}_m-t)}p_b(\widetilde{\tau}_m)b_{k(\widetilde{\tau}_m)}(a,\widetilde{\tau}_m) \\ +\int_t^{\widetilde{\tau}_m} e^{-r(s-t)}\left[u_{k(s)}(a,b_{k(s)}(a,s))-[rp_b(s)-\dot{p}_b(s)]\,b_{k(s)}(a,s)\right]ds\end{aligned}$$

Nash bargaining and pseudo arrival rate

At each meeting, intermediaries have bargaining power η. The Nash-bargaining problem faced by agents in state j at time t and with outstanding illiquid holding a is

$$[a_j(t),\varphi_j(a,t)] = \arg\max_{\{a'\geq 0,\varphi\geq 0\}} \{\varphi\,[a'-a]\}^{\eta} \times \left\{V_j(a',t)-V_j(a,t)-p_a(t)\,[a'-a]-p_b(t)\,[b_j(a')-b_j(a)]-\varphi\,[a'-a]\right\}^{1-\eta}$$

At a meeting, each investor (net buyer or net seller) takes into account the fact that a modification of its illiquid asset holdings will automatically trigger a readjustment of its portfolio position in the liquid asset. The optimal choice of illiquid holding follows

$$a_j(t) = \arg\max_{a'\geq 0}\left[V_j(a',t)-p_a(t)a'-p_b(t)b_j(a')\right].$$

While optimal fees derived from the first-order conditions are equal to a fraction η of the gains from trade,

$$\begin{aligned} TG_j(a,t) &= V_j(a_j(t),t)-V_j(a,t)-p_a(t)\left[a_j(t)-a\right]-p_b(t)\left[b_j(a_j(t),t)-b_j(a,t)\right] \\ \phi_j(a,t) &= \eta\; TG_j(a,t) \\ \varphi_j(a,t) &= \phi_j(a,t)/\left[a_j(t)-a\right] \end{aligned}$$

Plugging back in the optimal fees into the value function $V_i(a,t)$,

$$\mathbb{E}_t\left[\begin{array}{c} p_b(t)b_i(a,t)+\int_t^{\tilde{\tau}_m} e^{-r(s-t)}\hat{u}_{i(s)}(a,s)ds - e^{-r(\tilde{\tau}_m-t)}p_b(\tilde{\tau}_m)b_{i(\tilde{\tau}_m)}(a,\tilde{\tau}_m) \\ +e^{-r(\tilde{\tau}_m-t)}\left\{\begin{array}{c} (1-\eta)\max\limits_{\{a'\geq 0\}}\left[\begin{array}{c} V_{i(\tilde{\tau}_m)}(a',\tilde{\tau}_m)-p_a(\tilde{\tau}_m)\left(a'-a\right) \\ -p_b(\tilde{\tau}_m)\left[b_{i(\tilde{\tau}_m)}(a',\tilde{\tau}_m)-b_{i(\tilde{\tau}_m)}(a,\tilde{\tau}_m)\right]\end{array}\right] \\ +\eta V_{i(\tilde{\tau}_m)}(a,\tilde{\tau}_m)\end{array}\right\}\end{array}\right]$$

shows that asset holdings are modified at the pseudo arrival rate $\kappa \equiv \alpha(1-\eta)$. Hence, after simplification and denoting by τ_m the corresponding pseudo stopping time, the value function simplifies to

$$\mathbb{E}_t\left[\begin{array}{c} p_b(t)b_i(a,t)+\int_t^{\tau_m} e^{-r(s-t)}\hat{u}_{i(s)}(a,s)ds \\ +e^{-r(\tau_m-t)}\left\{p_a(\tau_m)a+\max\limits_{\{a'\geq 0\}}\left[V_{i(\tau_m)}(a',\tau_m)-p_a(\tau_m)a'-p_a(\tau_m)b_{i(\tau_m)}(a')\right]\right\}\end{array}\right].$$

Entering the market

Eventually, an investor entering the market at date t would optimally choose a' to maximize

$$\mathbb{E}_t\left[\int_t^{\tau_m} e^{-r(s-t)}\hat{u}_{i(s)}(a',s)ds\right]-\left[p_a(t)-\mathbb{E}_t[e^{-r(\tau_m-t)}p_a(\tau_m)]\right]a'.$$

And defining the modified utility[38] and holding costs

$$\begin{aligned}\widehat{U}_j(a,t) &= (r+\kappa)\mathbb{E}_t\left[\int_t^{\tau_m} e^{-r(s-t)}\widehat{u}_{i(s)}(a,s)ds\,|\,i(t)=j\right],\\ q_a(t) &= (r+\kappa)\left[p_a(t)-\kappa\int_0^{\infty} e^{-(r+\kappa)s}p_a(t+s)ds\right],\end{aligned}$$

the optimization can be expressed in the more compact form[39]:

$$\max_{a'\geq 0}\left[\widehat{U}_j(a')-q_a(t)a'\right].$$

Value function and OTC fees

We claim[40] that the value function can be expressed as

$$\begin{aligned}V_i(a,t) &= \frac{\widehat{U}_i(a,t)}{r+\kappa}+\mathbb{E}_t[e^{-r(s-\tau_m)}p_a(\tau_m)]a+p_b(t)b_i(a,t)+C_i(t)\\ &= \frac{\widehat{U}_i(a,t)}{r+\kappa}+\left[p_a(t)-\frac{q_a(t)}{r+\kappa}\right]a+p_b(t)b_i(a,t)+C_i(t)\end{aligned}$$

where

$$C_i(t)=\mathbb{E}_t\left[\sum_{n=1}^{\infty} e^{-r(t_{n_t+n}-t)}\left\{\frac{\widehat{U}_{k(t_{n_t+n})}(a_{i(t_{n_t+n})}(t_{n_t+n}),t_{n_t+n})}{r+\kappa}-\frac{q_a(t_{n_t+n})}{r+\kappa}a_{k(t_{n_t+n})}(t_{n_t+n})\right\}\right]$$

where n_t is the time t counting process for OTC meetings. The $C_i(t)$ function captures the future gains from trade beyond the next OTC meeting, and is independent of the current holdings in the illiquid asset.

Fees

From the Nash-bargaining result, OTC fees are defined as

$$\phi_i(a,t)=\eta\left[\frac{\widehat{U}_i(a_i(t),t)-\widehat{U}_i(a,t)}{r+\kappa}-\frac{q_a(t)}{r+\kappa}[a_i(t)-a]\right].$$

[38]Strict concavity is preserved since for any (i,a)

$$\begin{aligned}\frac{\partial^2}{\partial a^2}\widehat{u}_i(a) &= \frac{\partial}{\partial a}\partial_1 u_i(a,b_i(a))=\partial_{11}u_i(a,b_i(a))+\partial_{12}u_i(a,b_i(a))\times b_i'(a)\\ &= \partial_{11}u_i(a,b_i(a))-\frac{\partial_{12}u_i(a,b_i(a))^2}{\partial_{22}u_i(a,b_i(a))}=Hess\;\; u_i(a,b_i(a))/\partial_{22}u_i(a,b_i(a)).\end{aligned}$$

And the conclusion derives from the strictly positive Hessian and the strictly negative second-order derivative of the original utility flow function.

[39]Note that $\mathbb{E}_t[e^{-r(\tau_m-t)}p_a(\tau_m)]=\kappa\int_0^{\infty}e^{-(r+\kappa)s}p_a(t+s)ds$

[40]See Lagos and Rocheteau (2009), online appendix, for a similar derivation of value functions and trading fees under Poisson search with Nash bargaining.

E Proof of freeze demand impact

The outline of the proof for Lemma 10 provides some interesting insights. First, allocation-wise, the optimal choice of $b-$holdings, conditional on prices and asset a, follows the same program before and after the freeze. Indeed, in a post-freeze stationary environment with $\widetilde{p_b}$ the prevailing exchange price, trades only occur at type switches, and the maximization still proceeds for all $i \in \mathscr{I}$ as

$$b_i(a) = \arg\max_{b' \geq 0} \left[u_i(a, b') - r\widetilde{p_b}b' \right].$$

The element that is modified is the distribution of investors. In equilibrium, if the freeze occured from a stationary environment where for each $j \in \mathscr{I}$, a fraction π_j of the population held a quantity a_j of asset a, adopting similar notations to those in section (3.5), for all $(i,j) \in \mathscr{I}^2$, the measure of investors in preference state i and holding a_j is $\widetilde{M}_{j,i} = \pi_j \pi_i$.

At this point, one can compute the change in demand for type-b assets following the freeze, before any price adjustment as

$$\Delta D^b = \sum_{(i,j) \in \mathscr{I}^2} \left[\widetilde{M}_{j,i} - M_{j,i} \right] b_i(a_j) = \frac{\alpha}{\delta + \alpha} \left\{ Q^b - \sum_{j \in \mathscr{I}} \pi_j b_j(a_j) \right\};$$

which concludes the proof in a world with fixed supply Q^b.

F Proof of the existence of a unique equilibrium

After adapting to the fact that the OTC market is not Walrasian, the logic of the proof is somewhat similar to that found in Mas-Colell et al. (1995) exercise 10.G.1.

We define the optimization program of a planner who would take all frictions (search and bargaining) as given. As such, from the model derivations (and in particular the resulting properties of the Nash-bargaining round), investors can seen as having access to the competitive inter-dealer market at the realization times of a Poisson process of intensity κ.

The planner's program nests two optimizations. One allocates b-type assets at every point in time

$$\max_{\{b_i(a,s)\}_{i \in \mathscr{I}}} \int_0^\infty \sum_{i \in \mathscr{I}} M_i(a,s) u_i(a, b_i(a,s)) da$$
$$\text{such that} \int_0^\infty \sum_{i \in \mathscr{I}} M_i(a,s) b_i(a,s) da = Q^b(s).$$

It admits for Lagrangian

$$L^b(\{a\},s) = \int_0^\infty \sum_{i\in\mathcal{I}} M_i(a,s)u_i(a,b_i(a,s))\,da + \lambda^b(\{a\},s)\left[Q^b(s) - \int_0^\infty \sum_{i\in\mathcal{I}} M_i(a,s)b_i(a,s)\,da\right],$$

and for first-order condition

$$0 = \partial_2 u_i(a,b_i(a,s)) - \lambda^b(\{a\},s) \quad \forall i,a. \tag{31}$$

The other allocates a-type assets at every point in time among investors having contacted intermediaries. From the point of view of the planner, the benefit for an investor of holding a given quantity a of that asset is worth the net-utility flow $u_i(a,s) = u_i(a,b_i(a,s)) - \lambda^b(\{a\},s)b_i(a,s)$. Hence, optimizing holdings over the no-reallocation horizon for all agents boils down to

$$\max_{\{a_j(t)\}_{j\in\mathcal{I}}} (r+\kappa)\sum_{j\in\mathcal{I}} n_j(t)\mathbb{E}_t\left[\int_t^{\tau_m} e^{-r(s-t)}u_{i(s)}\left(a_j(t),s\right)ds\,|i(t)=j\right]$$
$$\text{such that } \sum_{j\in\mathcal{I}} n_j(t)a_j(t) = Q^a(t).$$

The Lagrangian of that program

$$L^a(t) = (r+\kappa)\sum_{j\in\mathcal{I}} n_j(t)\mathbb{E}_t\left[\int_t^{\tau_m} e^{-r(s-t)}u_{i(s)}\left(a_j(t),s\right)ds\,|i(t)=j\right] + \lambda^a(t)\left[Q^a(t) - \sum_{j\in\mathcal{I}} n_j(t)a_j(t)\right]$$

admits as first-order condition

$$0 = (r+\kappa)\mathbb{E}_t\left[\int_t^{\tau_m} e^{-r(s-t)}\partial_1 u_{i(s)}\left(a_j(t),b_i\left(a_j(t),s\right)\right)ds\,|i(t)=j\right] - \lambda^a(t) \quad \forall j \tag{32}$$

given the nested program envelope condition $\partial_a u_{i(s)}(a,s) = \partial_1 u_{i(s)}(a,b_{i(s)}(a))\ \forall\ s \geq t$.

These programs lead to the same first-order conditions as those that determine optimal allocations in a world without a central planner. Hence, I show next that the central planner allocations are uniquely defined, which guarantees the existence and uniqueness of prices as defined by the market clearing conditions.

The concavity of the nested program for asset b is immediate. We then show the concavity of the nesting program for asset-a by first differentiating (31) with respect to the two conditioning variables

$$[a] \quad : \quad 0 = \partial_{12}u_i(a,b_i(a,s)) + \partial_{22}u_i(a,b_i(a,s))\frac{\partial b_i(a,s)}{\partial a} - \frac{\partial \lambda^b(s)}{\partial a}$$
$$[\lambda^b] \quad : \quad 0 = \partial_{22}u_i(a,b_i(a,s))\frac{\partial b_i(a,s)}{\partial \lambda^b(s)} + 1$$

to derive that the second-order derivative of the planner-modified utility function

$$\begin{aligned}
\partial_{aa}u_{i(s)}(a,s) &= \partial_{11}u_{i(s)}(a,b_{i(s)}(a)) + \partial_{12}u_{i(s)}(a,b_{i(s)}(a)) \times \left[\frac{\partial b_{i(s)}(a)}{\partial a} + \frac{\partial b_{i(s)}(a)}{\partial \lambda^b(s)}\frac{\partial \lambda^b(s)}{\partial a}\right]. \\
&= \partial_{11}u_{i(s)}(a,b_{i(s)}(a)) + \frac{\partial_{12}u_{i(s)}(a,b_{i(s)}(a,s))}{\partial_{22}u_{i(s)}\left(a,b_{i(s)}(a,s)\right)}\left[\frac{\partial \lambda^b(s)}{\partial a} - \partial_{12}u_i(a,b_i(a,s)) - \frac{\partial \lambda^b(s)}{\partial a}\right] \\
&= \frac{Hess\ u_{i(s)}(a,b_{i(s)}(a))}{\partial_{22}u_{i(s)}(a,b_{i(s)}(a))}.
\end{aligned}$$

is always negative under the utility regularity conditions.

G Endogenous distribution of assets

Lemma 5 is stated as:

Suppose that the price paths $p_a(s)$ and $p_b(s)$ are differentiable, and that the initial condition $Q^b(t) = Q^b$ is such that $p_b(t) - p_a(t) \notin \left[-\Gamma\left(Q^b\right), \Upsilon\left(Q^b\right)\right]$. Then, a bounded path, $Q^b(s)$ solves the intermediaries' problem, if and only if it verifies the no-bubble condition

$$\lim_{T\to+\infty} e^{-rT}\left[p_b(T) - p_a(T)\right]Q^b(T) = 0,$$

together with the optimality condition for all $s > t$

$$r\left[p_b(s) - p_a(s) - \Upsilon\left(Q^b(s)\right)I_{\dot{Q}^b(s)>0} + \Gamma\left(Q^b(s)\right)I_{\dot{Q}^b(s)<0}\right] = \Xi(s)$$

whenever $\dot{Q}^b(s) \neq 0$; and where the dynamic part $\Xi(s)$ is defined by

$$\Xi(s) = \dot{p}_b(s) - \dot{p}_a(s) - \dot{Q}^b(s)\left[I_{\dot{Q}^b(s)>0}\Upsilon'\left(Q^b(s)\right) - I_{\dot{Q}^b(s)<0}\Gamma'\left(Q^b(s)\right)\right].$$

Proof. Let's define the transfer objective function, starting from $Q^b(t) = Q^b$

$$\digamma(t,T) = \int_t^T e^{-r(s-t)}\left[p_b(s) - p_a(s) - I_{\dot{Q}^b(s)>0}\Upsilon(Q^b(s)) + I_{\dot{Q}^b(s)<0}\Gamma(Q^b(s))\right]\dot{Q}^b(s)ds$$

Integrating it by part obtains

$$\mathcal{F}(t,T) = e^{-r(T-t)}\left[p_b(T)-p_a(T)-I_{\dot{Q}^b(T)>0}\Upsilon(Q^b(T))+I_{\dot{Q}^b(T)<0}\Gamma(Q^b(T))\right]Q^b(T)$$
$$-\left[p_b(t)-p_a(t)-I_{\dot{Q}^b(t)>0}\Upsilon(Q^b)+I_{\dot{Q}^b(s)<0}\Gamma(Q^b)\right]Q^b$$
$$+\int_t^T e^{-r(s-t)}r\left[p_b(s)-p_a(s)-I_{\dot{Q}^b(s)>0}\Upsilon(Q^b(s))+I_{\dot{Q}^b(s)<0}\Gamma(Q^b(s))\right]Q^b(s)ds$$
$$-\int_t^T e^{-r(s-t)}[\dot{p}_b(s)-\dot{p}_a(s)]Q^b(s)ds$$
$$-\int_t^T e^{-r(s-t)}\dot{Q}^b(s)\left[-I_{\dot{Q}^b(s)>0}\Upsilon'(Q^b(s))+I_{\dot{Q}^b(s)<0}\Gamma'(Q^b(s))\right]Q^b(s)ds$$
$$+\int_t^T e^{-r(s-t)}\delta\left(\dot{Q}^b(s)\right)\ddot{Q}^b(s)\left[\Upsilon(Q^b(s))+\Gamma(Q^b(s))\right]Q^b(s)ds$$

where $\delta(.)$ is the Dirac function. The no-bubble condition ensures that the first term tends to zero when T goes to infinity. And from the super contact condition $\ddot{Q}^b(s)=0$. So that, after identifying the dynamic term for all $s>t$,

$$\Xi(s)=\dot{p}_b(s)-\dot{p}_a(s)-\dot{Q}^b(s)\left[I_{\dot{Q}^b(s)>0}\Upsilon'(Q^b(s))-I_{\dot{Q}^b(s)<0}\Gamma'(Q^b(s))\right],$$

we express the objective function as

$$\mathcal{F}(t,\infty) = -\left[p_b(t)-p_a(t)-I_{\dot{Q}^b(t)>0}\Upsilon(Q^b)+I_{\dot{Q}^b(t)<0}\Gamma(Q^b)\right]Q^b$$
$$+\int_t^\infty e^{-r(s-t)}r\left[p_b(s)-p_a(s)-I_{\dot{Q}^b(s)>0}\Upsilon(Q^b(s))+I_{\dot{Q}^b(s)<0}\Gamma(Q^b(s))\right]Q^b(s)ds$$
$$-\int_t^\infty e^{-r(s-t)}\Xi(s)Q^b(s)ds.$$

The first term takes into account the outstanding stock $Q^b(t)=Q^b$. The other term leads to the result in lemma 5. The condition stated in the lemma is sufficient. It is necessary since, if it were not verified, intermediaries could have the incentive to transfer infinite quantities of assets across platforms. □

The following corollary details the dynamics of the three state variables (p_a, p_b, Q^b) presented in the phase diagram of Figure 5.

Corollary 11. *Given $Q^b(0)$, and $(p_a(Q^b(0)), p_b(Q^b(0)))$ the prices which would prevail at this level of assets, the equilibrium dynamics of $Q^b(t)$ follow one of three possible paths:*
(i) if $p_b(Q^b(0))-p_a(Q^b(0))\geq\Upsilon(Q^b(0))$, given demand and cost structure necessarily $D^b(0)>$

$Q^b(0)$*, and the dynamics are*

$$\begin{aligned} D^b(t) &= Q - Q^a(t), \\ \dot{Q}^a(t) &= \alpha\left[D^a(t) - Q^a(t)\right] \leq 0, \text{ and} \\ \dot{p}_b(s) - \dot{p}_a(s) - \dot{Q}^b(s)\Upsilon'(Q^b(s)) &= r\left[p_b(s) - p_a(s) - \Upsilon(Q^b(s))\right]. \end{aligned}$$

(ii) if $p_b(Q^b(0)) - p_a(Q^b(0)) \geq \Gamma(Q^b(0))$*, given demand and cost structure necessarily* $D^a(0) < Q^a(0)$*, and the dynamics are*

$$\begin{aligned} D^b(t) &= Q - Q^a(t), \\ \dot{Q}^a(t) &= \alpha\left[D^a(t) - Q^a(t)\right] \geq 0, \text{ and} \\ \dot{p}_b(s) - \dot{p}_a(s) + \dot{Q}^b(s)\Gamma'(Q^b(s)) &= r\left[p_b(s) - p_a(s) + \Gamma(Q^b(s))\right]. \end{aligned}$$

(iii) if $-\Gamma(Q^b(0)) > p_b(Q^b(0)) - p_a(Q^b(0)) > \Upsilon(Q^b(0))$*, i.e., between the 2 boundaries,*

$$\begin{aligned} D^b(t) &= Q - Q^a(t), \\ D^a(t) &= Q^a(t), \text{ and} \\ Q^a(t) &= Q^a(0). \end{aligned}$$

H Fees: OTC–CEX versus OTC-only environments

Figure A2 shows the decomposition of total fees captured by intermediaries in both the OTC–CEX and OTC-only environments. It identifies an OTC trading speed at which fee capture is maximized. This pattern results from two opposing forces: the intensive margin, represented by unit fees, increases with trading frictions, while the extensive margin, trading volume, decreases as frictions rise.

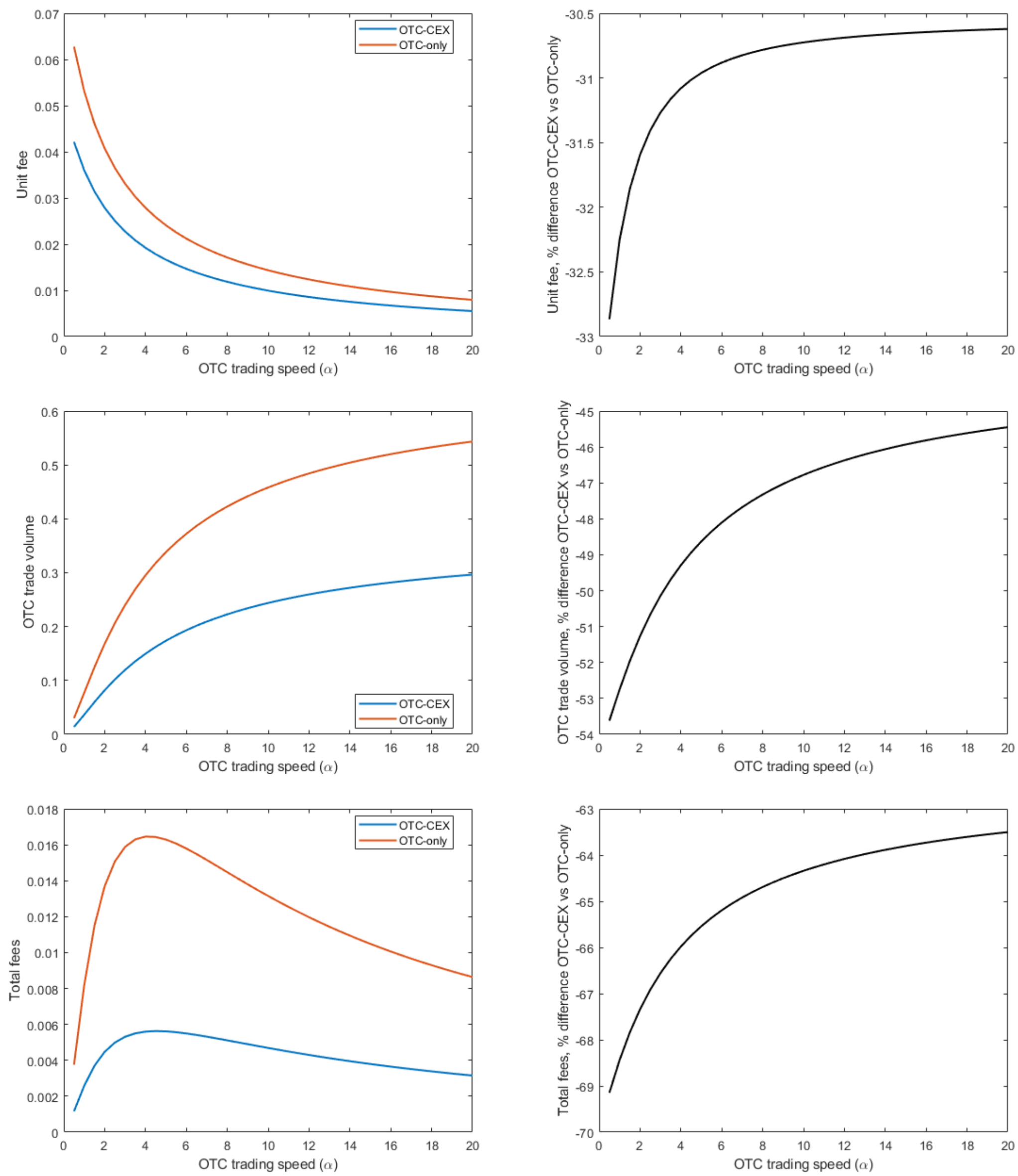


Figure A2: Comparison of trading fees in OTC–CEX versus OTC-only environments under different OTC trading frictions. The x-axis reports the OTC trading speed α. The first row displays unit fees charged by intermediaries. The second row shows trading volume. The third row shows total fees captured. The left column compares absolute levels under the OTC–CEX and OTC-only regimes. The right column reports their percentage difference.