Portfolio Choice and Settlement Frictions: A Theory of Endogenous Convenience Yields*

Javier Bianchi† and Saki Bigio‡

February 2026

Abstract

We study settlement frictions that arise from the need to finance negative balances through an over-the-counter (OTC) market. We derive a closed-form expression for the endogenous convenience yield and show how it can be incorporated into a canonical portfolio problem. Using this framework, we examine how shifts in settlement frictions affect liquidity premia, the volume of overnight funding, the dispersion of market rates, and optimal portfolio allocations. From a normative perspective, we show that in the competitive equilibrium, investors may either over- or under-invest in liquid assets; moreover, both higher risk aversion and tighter aggregate liquidity increase the likelihood of under-accumulation. Finally, we use our framework to discuss the implementation of monetary policy and argue that a transition from abundant to an ample regime need not be smooth.

Keywords: Asset Pricing, Portfolio Theory, OTC markets, convenience yields

JEL Classification: E51, E52, E58, G12, G21

*We thank Andy Atkeson, David Baqaee, Ricardo Lagos, and Pierre-Olivier Weill for helpful comments. Ali Haider Ismail, Seungyub Han, Jorge Mondragon, Luis Yépez, and Chenming Zhang provided invaluable research assistance. A previous version of this paper was circulated with the title "Settlement Frictions and Portfolio Theory." The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†Minneapolis Fed and NBER, email javier.i.bianchi@gmail.com

‡Department of Economics, UCLA and NBER, email sbigio@econ.ucla.edu

1. Introduction

The empirical finance literature has established that certain assets command substantial convenience yields—premia not accounted for by their cash flows alone (Krishnamurthy and Vissing-Jørgensen, 2012; Nagel, 2016). Convenience yields vary across assets and over time, shaping the transmission of monetary policy, fiscal capacity, and international capital flows. Yet most asset-pricing models treat them as exogenous, while existing microfounded approaches are typically less tractable. This paper develops a tractable microfoundation for convenience yields generated by trading frictions in over-the-counter (OTC) markets and embeds it in canonical portfolio theory.

In our theory, investors are subject to payoff risk, stemming from fundamental variations in returns, and liquidity risk. Liquidity risk arises because cash flows—such as deposit withdrawals for banks, margin calls for hedge funds, or claim payouts for insurance companies—are unpredictable and must be settled in an OTC market. Investors hold buffers of liquid assets to meet settlement obligations. When liquidity needs exceed available buffers, investors must borrow cash in a frictional OTC market. The interaction between portfolio-driven cash needs and OTC market frictions generates an endogenous convenience yield that lowers the returns on liquid assets. When cash needs correlate with asset returns, they generate an additional liquidity risk premium beyond the standard risk premium. Thus, convenience yields reflect both a first-order effect and the covariance between cash needs and asset payoffs. Crucially, because these yields depend endogenously on investors' portfolios and OTC-market conditions, they vary with policy, market tightness, and trading efficiency.

Our analysis yields a tractable derivation of convenience yields, which can be easily embedded in standard portfolio problems. To do so, we build on the sequential OTC framework of Afonso and Lagos (2015b), with two key innovations. First, instead of taking settlement positions as given, here they emerge endogenously from portfolio choices. Second, we model investors who delegate settlement trades to many small traders, following Shi (1997) and the OTC model of Atkeson, Eisfeldt and Weill (2015). Taking limits as trader size approaches zero, we show that the OTC rates can be solved independently of the equity of the investors' wealth. Furthermore, taking the continuous time limit yields closed-form solutions of the entire equilibrium path of trades and rates.

Settlement frictions manifest as a piecewise linear convenience yield function that maps a settlement position that each investor needs to close into a marginal cost of liquidity. This function is kinked at zero, reflecting the asymmetry between borrowing costs (for investors with a cash deficit) and lending returns (for investors with a cash surplus). The slopes of the convenience-yield function are equilibrium objects that depend on bargaining power, the market tightness, and the matching technology. The asymmetry in the convenience yield induces risk-averse behavior even under risk-

neutral preferences and implies that assets that generate volatile cash needs command higher convenience yields. Moreover, when cash needs covary with asset returns—as when margin calls intensify in downturns—an additional liquidity risk premium emerges. Thus, convenience yields reflect both pure OTC frictions and the interaction between liquidity risk and return risk.

Our analysis reveals several properties linking OTC markets to convenience yields. First, the type of matching technology shapes market dynamics. While under Cobb-Douglas matching, one side of the market can be depleted before the trading session is over, permitting an efficient allocation of funds, under Leontief matching, trading remains active at all times. Thus, the Cobb-Douglas case showcases that convenience yields can be zero even when the OTC market exhibits some volume. For other matching functions, convenience yields are zero only if there is no OTC market volume and investors are satiated with cash.[1]

These findings have important implications for monetary policy implementation, and in particular highlight that the transition from an "abundant" to a merely "ample" reserve regime can be nonlinear, with relatively small changes in frictions or aggregate liquidity producing outsized changes in overnight rate dispersion and funding conditions—an important consideration for the implementation and monitoring of monetary policy.

In addition, we show that while market tightness unambiguously increases convenience yields by steepening the convenience yield function, matching efficiency may have non-monotonic effects. The intuition behind this non-monotonicity—perhaps a surprising result—is that higher matching efficiency benefits the short side of the market, raising convenience yields when liquidity is scarce but potentially lowering them when it is abundant. Furthermore, our closed-form solutions facilitate the identification of structural OTC market parameters and shocks from observable data for trading volumes and spreads.

Finally, we show that the interaction between portfolio choice and settlement frictions gives rise to a pecuniary externality that leads to inefficient investment in liquid assets. Individual investors, when choosing portfolios, fail to internalize how their cash surpluses and deficits affect aggregate market tightness and the resulting congestion in the OTC market, which in turn determines the convenience yields faced by all participants. This congestion externality drives a wedge between the private and social marginal returns to holding liquid assets. We characterize the conditions under which competitive equilibria feature over- or under-investment in liquidity. Under risk neutrality, there is under-investment when aggregate cash has a stronger marginal impact on borrowing costs than on lending returns—a situation that arises when investors in surplus outnumber those in deficit—and over-investment in the opposite case. When markets are balanced, the competitive

[1]Our analysis shows that only certain matching functions, those with infinite rate of decay, can generate near-zero convenience yields with positive trading volumes.

equilibrium is efficient under risk neutrality but becomes inefficient under risk aversion, leading to under-investment in liquidity. These findings have implications for liquidity regulation, indicating that optimal policy depends critically on market conditions.

Literature Review. Our paper is related to a large literature on portfolio choice and asset pricing with liquidity frictions, such as portfolio constraints or transaction costs. Important examples include Constantinides (1986), Basak and Cuoco (1998), Vayanos and Vila (1999), Constantinides and Duffie (1996), Krueger and Lustig (2010), Holmstrom and Tirole (2001), Acharya and Pedersen (2005), Lagos (2010). These studies typically treat liquidity frictions as exogenous. Our goal in this paper is to develop a tractable framework where liquidity premia emerge endogenously and explore the implications for asset portfolios.

Following Duffie, Garleanu and Pedersen (2005), a burgeoning literature on OTC markets has studied environments where assets are traded in the presence of search frictions.[2] This literature has identified how features of the trading environment, such as the speed of transactions and the heterogeneity in the motives for trade, affect trading volumes and impact liquidity premia. While the literature began with strong restrictions on portfolio holdings— namely, binary holdings—work by Lagos and Rocheteau (2009) allows for arbitrary portfolio holdings, bringing this literature closer to standard portfolio theory. Hugonnier, Lester and Weill (2022) consider heterogeneity in asset selling speeds. Uslu (2019) introduces risk-averse behavior into this class of models. He shows that trading speed affects asset values because time is discounted, a feature that depresses the option value of selling assets when gains from trade emerge. He and Milbradt (2014) study how liquidity in an OTC secondary market for bonds affects the issuer's default risk. Silva, Passadore and Kargar (2023) study portfolio problems that explicitly take into account trading times when agents want to modify their portfolios.[3] In the language of Hugonnier, Lester and Weill (2025), these papers study a semi-centralized setup, while we study a purely decentralized setup. Moreover, we focus on OTC frictions in the market for borrowing and lending, as opposed to the sale of assets.[4]

As stated above, the sequential OTC market is inherited from Afonso and Lagos (2015b). In a further application, Afonso and Lagos (2015a) obtain closed-form formulas for volumes and OTC rates that mirror those found here. Afonso and Lagos (2015a) restrict positions to binary holdings and assume an increasing returns-to-scale matching function. Here, we show that, by working with

[2]This literature has run in parallel to the New Monetarist literature, pioneered by Kiyotaki and Wright (1993). See Williamson and Wright (2010); Lagos, Rocheteau and Wright (2017) for recent surveys.

[3]In our model, time discounting plays no role, as trading occurs within a single period, although the sequence of trades matters because it affects the terms of trade when two investors match.

[4]In the work of Lagos and Wright (2005), the traded object corresponds to goods, and money facilitates transactions. Another related study is Lester, Postlewaite and Wright (2012) who considers a model where the recognizability of the asset determines whether they are used for transactions.

a large number of traders within each investor—as in Shi (1997) and Atkeson, Eisfeldt and Weill (2015)—we can readily embed the OTC block into portfolio theory. We furthermore characterize outcomes for a more general class of matching processes. In addition, we provide comparative statics with respect to market tightness and matching efficiency, show how these outcomes can be mapped to observable empirical moments, and conduct a normative analysis of investors' portfolio choices.

Our normative analysis is related to a broad literature on the welfare properties of competitive equilibria with financial frictions—in particular, those studies analyzing the efficiency of risk-taking decisions. One branch of the literature studies over- or under-investment in liquid assets in Walrasian markets where the risk-free rate affects the degree of enforcement and risk sharing (e.g., Jacklin, 1987; Bhattacharya and Gale, 1987; Farhi, Golosov and Tsyvinski, 2009; Yared, 2013; Geanakoplos and Walsh, 2017). In contrast, we consider a setting with an OTC market where a pecuniary externality emerges from congestion in the interbank market. The externality we study is related to other congestion externalities in the search and matching literature. In particular, Uslu (2019) identifies an externality where fast investors can capture a private transaction surplus larger than their contribution to surplus creation.[5] In addition, Wong and Zhang (2023) shows that the degree of search and intermediation is inefficient in the OTC market where search intensity is endogenous. Different from these contributions, our approach compares the portfolio choices of individual investors vis-à-vis a social planner that takes as given the financial arrangements in the OTC market. Our analysis is closest to Arseneau, Rappoport and Vardoulakis (2020), who consider a model where firms issue long-term bonds that are retraded by risk-neutral investors in OTC markets and show that liquidity provision is constrained inefficient. Apart from differences in our framework, a contribution of our paper is to show that risk aversion plays a central role in determining the direction of the inefficiency.[6]

Finally, our earlier work (Bianchi and Bigio, 2022) used the convenience-yield function derived here to analyze the transmission and implementation of monetary policy through the banking system; this paper develops the theoretical foundations.[7] Related studies on liquidity management of financial intermediaries include Arce, Nuño, Thaler and Thomas (2019) and Bigio and Sannikov (2019) on optimal reserve policy in models with nominal rigidities and incomplete markets, respectively; Lenel, Piazzesi and Schneider (2019) and Piazzesi, Rogers and Schneider (2019) on the short-rate disconnect and on monetary transmission with nominal rigidities, respectively; Bianchi, Bigio and

[5] In contrast, in a model with exogenous search effort, Afonso and Lagos (2015a,b) find that the competitive equilibrium coincides with the efficient allocation.

[6] A different strand of the literature analyzes pecuniary externalities that can result in excessive leverage (see e.g. Caballero and Krishnamurthy, 2001; Lorenzoni, 2008; Bianchi, 2011; Dávila and Korinek, 2018; Amador and Bianchi, 2024).

[7] In addition, we analyze the efficiency of portfolio choices.

Engel (2023) on exchange-rate determination in an international context; Lopez-Salido and Vissing-Jorgenson (2023) on quantitative tightening; Bittner, Jamilov and Saidi (2025) on assortative matching in the German interbank market using a similar framework; De Fiore, Hoerova and Uhlig (2018) on the role of collateral and interbank market freezes; and Ríos-Rull, Ritto, Takamura and Terajima (2024) on the credit channel in a model with productive externalities.

Roadmap. Section 2 presents the environment. Section 3 presents the main theoretical results. Section 4 presents the applications. Section 5 concludes.

2. Environment

We present an infinite-horizon model comprising a unit mass of investors who take portfolio positions given exogenous asset returns. Assets and liabilities differ in their payoffs across states and, crucially, in their settlement risks. This feature gives rise to a liquidity-management problem. Investors are subject to idiosyncratic shocks, and trade overnight loans in an over-the-counter (OTC) market.

Asset structure. There is a collection of assets, $\{a_i\}$, where $i \in \mathbb{I}=\{1, 2, ..., I\}$, where an investor takes a positive position when long and a negative position when short. Assets are real and differ in their realized returns, as well as their settlement properties. In addition, there is a special asset, m, which is used to settle all payoffs and represents "cash." Investors may hold a negative position in m during the period but must end each period with a positive cash position.

2.1 The Portfolio and Balancing Stages

Each period consists of two stages: a portfolio stage and a balancing stage. In the portfolio stage, investors take portfolio positions, given possible asset returns and cash-flow shocks. In the balancing stage, investors are subject to cash-flow shocks and may face a shortage or surplus of cash. When the investor has a cash deficit, it can borrow cash from another investor or through a backstop lender. When it has a surplus, it can lend to another investor or hold cash and earn a cash rate. We now describe each stage.

Portfolio Stage. In this first sub-stage, the individual investor makes portfolio holding decisions. The investor enters the period with a given initial amount of wealth or equity e_t. Equity is composed

of a portfolio of assets, $\{a_t^i\}_{i\in\mathbb{I}}$, cash, m_t, loans borrowed from other investors, f_t (which is negative when loans are provided to other investors), and borrowing from a backstop lender, w_t. That is,

$$e_t \equiv \sum_{i\in\mathbb{I}} R_t^i a_t^i + R_t^m m_t - \overline{R}_t^f f_t - R_t^w w_t, \tag{1}$$

where R_t denotes the realized gross return from period $t-1$ to t, with r_t denoting the corresponding *net* rates of return. We assume that $R_t^w \geq R_t^m$, which captures a penalty for emergency borrowing.[8] The return $\overline{R}_t^f$ is a weighted average rate of OTC loans, to be described below. This is the only rate of return determined endogenously.

Starting from a given equity level, the investor chooses the dividend—interpreted as an equity injection if negative—denoted by c_t, along with a new portfolio $\left\{\left\{\tilde{a}_{t+1}^i\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}\right\}$. We use *tildes* to denote a portfolio variable chosen at the portfolio stage. The budget constraint is

$$c_t + \sum_{i\in\mathbb{I}} \tilde{a}_{t+1}^i + \tilde{m}_{t+1} \leq e_t. \tag{2}$$

Balancing stage. Investors enter the balancing stage with the portfolio $\left\{\left\{\tilde{a}_{t+1}^i\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}\right\}$. At the beginning of the balancing stage, the investor experiences cash-flow shocks, $\{\omega_t^i\}$, to each of the assets in its portfolio

$$a_{t+1}^i = \tilde{a}_{t+1}^i(1-\omega_t^i),$$

where $\omega^i \in [-\infty, 1)$ follows a c.d.f. $\Phi^i(\cdot)$. If $\tilde{a}_{t+1}^i > 0$, a positive ω^i shock captures a redemption of assets—for instance, the early repayment of a loan or a callable bond—while a negative ω^i shock indicates a sudden increase in assets, such as the drawing of a credit line extended by the investor. Conversely, if $\tilde{a}_{t+1}^i < 0$, a positive ω^i shock represents a sudden withdrawal of liabilities, whereas a negative ω^i shock corresponds to an unexpected inflow of liabilities.

Cash-flow shocks are settled with cash. After the realization of cash-flow shocks, the investor's *cash surplus*

$$s\left(\left\{\tilde{a}_{t+1}^i\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}, \left\{\omega_t^i\right\}_{i\in\mathbb{I}}\right) = \underbrace{\tilde{m}_{t+1} + \sum_{i\in\mathbb{I}} \frac{R_{t+1}^i}{R_{t+1}^m}\omega_t \tilde{a}_{t+1}^i}_{\text{settlement balance}}. \tag{3}$$

The surplus s depends on the portfolio chosen in the portfolio stage, $\left\{\left\{\tilde{a}_{t+1}^i\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}\right\}$, and the vector of cash-flow shocks. When $\tilde{a}_{t+1}^i \cdot \omega^i > 0$, this shock represents a *positive cash-flow shock*; conversely, when $\tilde{a}_{t+1}^i \cdot \omega^i < 0$, it constitutes a *negative cash-flow shock*. Equation (3) shows that

[8] This assumption is consistent with monetary policy implementation, where the penalty rate is strictly above the interest on reserve balances.

if the investor faces sufficiently large negative cash-flow shocks, it may end up with a *cash deficit*. Implicit in this cash-flow accounting is the assumption that when settlement occurs, the return on each asset accrues to its holder, which is taken care of by the ratio of returns in (3).

If an investor ends up with a cash deficit, it must either borrow in the OTC market or resort to penalty borrowing. Conversely, if it has a surplus, it may lend in the OTC market or retain the cash. At the end of the period, cash holdings are given by:

$$m_{t+1} = s\left(\left\{\tilde{a}^i_{t+1}\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}, \left\{\omega^i_t\right\}_{i\in\mathbb{I}}\right) + f_{t+1} + w_{t+1} \geq 0. \tag{4}$$

The condition $m_{t+1} \geq 0$, combined with the assumption that $R^w \geq R^m$, implies that running a cash deficit is costly. Moreover, since investors may end up with different cash positions, there are gains from trade in the balancing stage. We assume that this money market is frictional, and operates as an *over-the-counter* (OTC) market—a natural assumption given that in practice, the money market is often decentralized: participants must search for counterparties with the opposite position. Because this market is OTC, access to funding is not guaranteed. As a result, funding deficits may be only partially covered through the OTC market, with the remainder financed via a backstop lender. Conversely, investors with a positive surplus may ultimately end up holding excess cash.

Given a cash-position s, investors first try to transact with another investor and then resort to a backstop lender or the cash market for any remainder deficit or surplus. We denote by $\left\{\Psi^+_t, \Psi^-_t\right\}$ the fractions of positive and negative surplus, respectively, that investors trade in the OTC market. That is, we have that

$$f_{t+1} = \begin{cases} -\Psi^-_t s_t & \text{if } s_t \leq 0, \\ -\Psi^+_t s_t & \text{if } s_t > 0, \end{cases} \qquad \text{and} \qquad w_{t+1} = \begin{cases} -(1-\Psi^-_t)s_t & \text{if } s_t \leq 0, \\ 0 & \text{if } s_t > 0. \end{cases} \tag{5}$$

Substituting (5) into (4), we can write the equity at $t+1$ as:

$$e_{t+1} \equiv \underbrace{\sum_{i\in\mathbb{I}} R^i_{t+1}\tilde{a}^i_{t+1} + R^m_{t+1}\tilde{m}_{t+1}}_{\text{portfolio return}} + \underbrace{\chi_{t+1}\left(s\left(\left\{\tilde{a}^i_{t+1}\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}\right)\right)}_{\text{convenience yield}}, \tag{6}$$

where χ_t represents the convenience-yield function

$$\chi_t(s) = \begin{cases} \chi^-_t s & \text{if } s \leq 0 \\ \chi^+_t s & \text{if } s > 0 \end{cases}, \tag{7}$$

where χ^- and χ^+ are given by

$$\chi_t^- = (\overline{R}_t^f - R_t^m)\Psi_{t-1}^- + (R_t^w - R_t^m)(1 - \Psi_{t-1}^-), \qquad \chi_t^+ = \Psi_{t-1}^+(\overline{R}_t^f - R_t^m).$$

Notice that the convenience yield function $\chi(s)$ is asymmetric: settlement deficits generate steep marginal losses, while surpluses generate lower marginal gains. This asymmetry plays an important role in the portfolio analysis.

Before we describe how the OTC market operates, we discuss several examples of assets that expose investors to liquidity risk, in addition to return risk.

2.2 Interpretation and Asset-Class Examples

One interpretation of the environment is that investors are banks, in which case "cash" corresponds to central bank reserves and backstop lending corresponds to the central bank's discount window. For non-bank financial institutions, cash typically takes the form of bank deposits, and liquidity backstops are pre-negotiated credit lines with banks rather than with the central bank. The OTC market in our model may correspond to the secured interbank market for banks or to the bilateral repo market for non-bank institutions, although we abstract from the possibility of default and the role of collateral here. In international contexts, cash corresponds to deposits held in the U.S. correspondent banking system, liquidity support to contingent credit lines with global banks, and the OTC market to short-term offshore funding markets such as the LIBOR market (see Bianchi et al., 2023).

We illustrate more concretely how cash needs arise across different asset classes.

(i) *Funding Outflows:* A natural interpretation of our framework arises in the context of institutions exposed to unpredictable funding outflows. One example is a bank whose liabilities consist of demand deposits, d, which are subject to withdrawal shocks and potentially a reserve requirement, $\rho \geq 0$, as in Bianchi and Bigio (2022). The bank holds a portfolio composed of a fully liquid asset and a fully illiquid asset. This case maps directly into our general framework, with $d = -\tilde{a}_t^1$ and the settlement position defined as

$$s(d, m, \omega) \equiv m - d\left(\frac{R^d}{R^m}\omega + \rho(1-\omega)\right).$$

The exact formulation applies to open-ended investment funds, where a redemption shock is analogous to a deposit withdrawal.

(ii) *Credit Lines:* A credit line specifies a multiple ω of an existing loan ℓ to be drawn upon the realization of a future state, at a pre-specified rate. The institution must then settle the exposure by delivering cash in the amount: $s(\ell, m, \omega) \equiv m - \omega\ell$.

(iii) *Margin Calls:* A margin call requires that, in all states where the future asset return falls below a threshold k, the borrower must post additional collateral. If the loan amount is b, the required cash settlement is

$$s(b, m) \equiv m - \max\left\{k - R^b_{t+1}, 0\right\} b.$$

In the next section, we examine the workings of the OTC market.

3. OTC Market

This section derives the key equilibrium objects in the OTC market, $\{\Psi^+_t, \Psi^-_t, \overline{R}^f_{t+1}\}$, that give rise to the convenience yield function (7). In general, these objects depend on the entire distribution of cash positions. As we show below, however, in our environment, they depend only on market tightness.

Definition 1. *We define market tightness θ as the ratio of total negative to total positive positions in the OTC market:*

$$\theta \equiv \frac{S^-}{S^+}, \quad \textit{where} \quad S^- = -\int_0^1 \min\{s^j, 0\}\, dj, \quad S^+ = \int_0^1 \max\{s^j, 0\}\, dj.$$

3.1 Two-sided matching

The market structure consists of a sequence of over-the-counter (OTC) markets, following Afonso and Lagos (2015b). Unlike that model, where investors trade directly, we assume that each investor delegates its trade to a large number of traders, each handling an order of size Δ. Traders close cash deficits and lend surpluses on behalf of investors.

Traders are segmented into borrowers and lenders, giving rise to a two-sided matching structure. If $\theta > 1$, the short side is the side with surplus. If $\theta < 1$, the short side is the side with a deficit.

Trading occurs over N rounds, indexed by $n \in \mathcal{N} \equiv 1, 2, \ldots, N$ over a fixed time interval—normalized to one. In each round, matches are formed according to a matching function. Once matched, borrowers and lenders bargain over the loan rate. Upon trade, the surplus trader transfers Δ cash to the deficit trader, which the corresponding investor repays at the beginning of the next portfolio stage. If a trader does not match in a given round, she may or may not match in the subse-

quent round. A trader in deficit who remains unmatched after round N borrows at the penalty rate R^w. Traders in surplus that are unmatched by round N, obtain the cash rate R^m.

We derive our analytical results in the limiting case where the order size Δ approaches zero, building on Shi (1997) and Atkeson et al. (2015), and the number of trading rounds N tends to infinity—i.e., the time within a trading round tends to zero. This limit is convenient because it avoids the combinatorial challenges that arise when the identity of the matches matters. Here, the terms of trade vary by round, depending on evolving surplus and deficit positions, but are independent of the identities of the counterparties.[9]

Matching function. Denote $\{S_n^+, S_n^-\}$ the aggregate mass of surplus and deficit positions after the realization of matches in round n and where $S_0^+ = S^+$and $S_0^- = S^-$. The number of matches in round $n+1$, z_{n+1}, is given by

$$z_{n+1} \equiv \lambda_N G(S_n^+, S_n^-), \quad n \in [0, 1, ..., N-1],$$

where λ_N is a parameter capturing the matching efficiency, indexed by the total number of rounds, N and $G : \mathbb{R}_+^2 \to \mathbb{R}_+$ is the matching function. The matching function satisfies the following properties.

Assumption 1 (Matching Function)**.** *The matching function G satisfies:*

i. No disposal: $G(0, \cdot) = G(\cdot, 0) = 0$.

ii. Constant returns to scale: $G(\cdot, \cdot)$ is homogeneous of degree one.

iii. Symmetry: $G(a, b) = G(b, a)$, $\forall a, b$.

iv. Weak exhaustion: $\lambda_N G(S_n^+, S_n^-) \leq \min\{S_n^+, S_n^-\}$.

v. Strictly increasing: $G_a(a, b) \geq 0, G_b(a, b) \geq 0$.

[9]The idea behind the modeling of fixed-sized order convention is that if a lender bank places lending orders that exceed its excess reserves, there is a chance it will not be able to have the funds to transfer to the borrowing bank. If a lending bank lacks the funds to transfer to the bank in deficit, the bank would violate a contract and face a large legal default cost. For sufficiently high costs, no bank will ever place lending orders above the amounts they hold. Assuming that orders are of a fixed size, investors cannot place more than $\eta(s, \Delta) \equiv \lfloor |s| / \Delta \rfloor$ orders. Here, $\lfloor x \rfloor$ is the floor function understood as the largest integer not greater than x. Because investors can only place integer numbers of orders, typically, there will be a remainder of reserves that cannot be lent or borrowed at the OTC market. These residuals can be borrowed (lent) at the penalty rate r^w (at the cash rate r^m) directly. Mathematically, this residual is $\phi(x, \Delta) = s - \lfloor |s| / \Delta \rfloor \Delta$.

No disposal implies that counterparts are needed for a match. Constant returns to scale imply that the number of matches scales proportionally with the size of the market. Symmetry assumes that the number of matches depends on the relative scarcity of borrowers or lenders, independently of which side is shorter. Weak exhaustion means that there cannot be more matches than the shortest side of the market. Monotonicity in each argument implies that adding more participants to either side results in more matches. We normalize $G(1,1) = 1$, without loss of generality, because we can rescale the number of matches by λ_N.

Finally, we note that it may be realistic to assume diminishing returns in the number of matches as a function of one side of the market, e.g., *weak concavity* $G_{aa}(a,b) \leq 0, G_{bb}(a,b) \leq 0$. The results that follow do not require this property, but it may be a desirable property in many applications.

3.2 Equilibrium

Given $\left\{S_0^+, S_0^-\right\}$, we construct the sequence of matches for each round, z_n, by tracking the evolution of surplus and deficit positions as follows:

$$S_n^+ \equiv S_{n-1}^+ - z_n \qquad \text{and} \qquad S_n^- \equiv S_{n-1}^- - z_n, \qquad \forall n \in \mathcal{N}.$$

As matches take place, the surplus and deficit positions shrink in the following round.[10] The recursion assumes implicitly that all matches result in trade. Accordingly, we define market tightness in round n as

$$\theta_n \equiv \frac{S_n^-}{S_n^+}, \qquad n \in \{0, 1, 2, \ldots, N\}.$$

and matching probabilities per round $\{\psi_n^+, \psi_n^-\}$ for a trader in surplus and deficit respectively as

$$\psi_n^+ \equiv \frac{z_n}{S_{n-1}^+} = \lambda_N G\left(1, \theta_{n-1}\right), \quad \psi_n^- \equiv \frac{z_n}{S_{n-1}^-} = \lambda_N G\left(\frac{1}{\theta_{n-1}}, 1\right) \qquad n \in \mathcal{N},$$

By convention, $\psi_{N+1}^+ = \psi_{N+1}^- = 0$, given that the last trading round with matches is N. By constant returns to scale, $\psi_n^+ = \theta_{n-1}\psi_n^-$.

Value functions. At each match, traders bargain over a loan rate, R_n^f, determined through Nash bargaining. The outside options for borrowers and lenders vary by round. In a match at the N-th round, the lender's outside option is R^m and the borrower's outside option is the outside borrowing rate, R^w. In earlier rounds, the outside option is the expected value of entering the next round with

[10] By no exhaustion and no disposal, if at any round either side of the market is exhausted—i.e., ends at zero—no further matches are formed.

an unmatched position.

Now, consider an individual trader who bargains on behalf of investor j. The trader is responsible for closing the order of size Δ. When bargaining, traders must form an expectation of their investor's equity position in each round of the OTC market. The equity position is random and depends on all other trades delegated by the investor, which involves a complex combinatorial problem. However, as we will see, taking the limit as $\Delta \to 0$ circumvents that challenge. The idea is that by taking this limit, we can resort to the law of large numbers, which renders OTC rates independent of the equity of the investors who in a match.

Exploiting the law of large numbers, if investor j has a cash deficit, $s^j < 0$, the trader forecasts that the fraction $\Psi^- \cdot (s^j + \Delta)$ will be borrowed across all rounds at an average rate $\overline{R}^f$ and $(1 - \Psi^-) \cdot (s^j + \Delta)$ will be borrowed at the penalty rate. If the investor has a cash surplus, $s^j > 0$, the trader expects that $\Psi^+ (s^j - \Delta)$ funds will be lent in the OTC market and that $(1 - \Psi^+) (s^j - \Delta)$ of the funds will be invested at the cash rate. This forecast is rationalized below when we take the size $\Delta \to 0$. Using these results together with (6), we have that the trader's forecast of the investor's equity is given by

$$e^j_{\Delta,n} = \underbrace{\sum_{i \in \mathbb{I}} a^i_{t+1} R^i_{t+1} + m_{t+1} R^m_{t+1} + \chi_{t+1}(s^j - \text{sign}(s^j)\Delta)}_{\equiv \mathcal{E}^j_{t+1}(\Delta)} + \text{sign}(s^j) \left(R^f_n - R^m_{t+1} \right) \Delta.$$

where the term $\mathcal{E}^j_{t+1}(\Delta)$ denotes the trader's forecast of equity, excluding its own.[11]

We let $V(e)$ denote the value function of the investor for a given level of initial equity which we take to be increasing and differentiable.[12] For convenience, we suppress the time subscript, since we will focus on the matching within the period. In addition, we let $J^+_M(n;\Delta)$ and $J^-_M(n;\Delta)$ denote the value functions of trader j in round n when in surplus and deficit, respectively. Also, for convenience, we suppress the dependence of $J^\pm_M$ on j, understanding that these refer to the values of matching for traders with a given level of equity and a surplus position. These values satisfy

$$J^{\text{sign}(s^j)}_M(n;\Delta) = V\left(\mathcal{E}^j(\Delta) + \text{sign}(s^j)\Delta \left(R^f_n - R^m\right)\right).$$

Likewise, we let $J^+_U(n;\Delta)$ and $J^-_U(n;\Delta)$ denote the corresponding unmatched trader's values. Re-

[11] For a discrete value $\Delta > 0$, investor's equity is a random variable that depends on the realization of trade outcomes across rounds.

[12] In section 4 this value function emerges from an investor's problem with concave utility. As well, the formulation allows investors to inherit a stochastic discount factor if one does not want to endow them with preferences.

cursively, these values are given by

$$J_U^{\text{sign}(s^j)}(n;\Delta) = \mathbb{E}\left[\psi_{n+1}^{\text{sign}(s^j)} J_M^j(n+1;\Delta) + \left(1-\psi_{n+1}^{\text{sign}(s^j)}\right) J_U^{\text{sign}(s^j))}(n+1;\Delta)\right].$$

The value of an unmatched trader is the probability of matching in the subsequent round times the value of matching its position, $\psi_{n+1}^{\text{sign}(s^j)} J_M^j(n+1;\Delta)$, plus the probability of not matching times the value of going to the subsequent round with an unmatched position. If by round N the trader is unmatched, the values are given by

$$J_U^+(N;\Delta) \equiv V\left(\mathcal{E}^j(\Delta)\right), \quad J_U^-(N;\Delta) \equiv V\left(\mathcal{E}^j(\Delta) - (R^w - R^m)\Delta\right).$$

Bargaining problem. With these value functions, we can describe the bargaining problem. Let us refer to the trader associated with a deficit position by j and the trader associated with a surplus position by k. Upon a match, an OTC rate solves the following Nash-bargaining problem:

$$r_n^f(\Delta) = \arg\max_{r_n^f} \left\{\mathcal{S}_n^-(\Delta)^\eta \times \mathcal{S}_n^+(\Delta)^{1-\eta}\right\}$$

where

$$\mathcal{S}_n^-(\Delta) = V\left(\mathcal{E}^j(\Delta) - \left(R_n^f - R^m\right)\Delta\right) - J_U^-(n;\Delta),$$
$$\mathcal{S}_n^+(\Delta) = V\left(\mathcal{E}^k(\Delta) + \left(R_n^f - R^m\right)\Delta\right) - J_U^+(n;\Delta).$$

3.3 Limit with Infinitesimal Orders

We now consider the limiting equilibrium as the order size vanishes, $\Delta \to 0$.[13] We have the following proposition:

Proposition 1 (Limit of Infinitesimal Trade Bargaining Problems). *Let $\Delta \to 0$. For all $n = 1, 2...$, the OTC rate r_n^f solves*

$$r_n^f \in \arg\max_{r \in [r^m + \chi_n^+,\, r^m + \chi_n^-]} \left(\chi_n^- - (r - r^m)\right)^\eta \left((r - r^m) - \chi_n^+\right)^{1-\eta},$$

[13] For ease of notation, here we work from now on with net interest rates.

where χ_n^+ *and* χ_n^- *satisfy*

$$\chi_n^+ = (r_{n+1}^f - r^m)\psi_{n+1}^+ + \chi_{n+1}^+(1-\psi_{n+1}^+), \tag{8}$$

$$\chi_n^- = (r_{n+1}^f - r^m)\psi_{n+1}^- + \chi_{n+1}^-(1-\psi_{n+1}^-), \tag{9}$$

with terminal conditions $\chi_N^+ = 0$ *and* $\chi_N^- = r^w - r^m$.

In addition, the solution yields

$$r_n^f = r^m + (1-\eta)\chi_n^- + \eta\chi_n^+, \qquad n = 1, 2, \ldots, N.$$

The proposition implies that the equilibrium OTC rate in round n is obtained by solving a system of linear difference equations. The rate equals the sum of the cash rate plus the weighted average of the outside options, $\chi_n^\pm$, and it is independent of the size of the surplus of the traders who meet. Importantly, because of the dynamic nature of the OTC market, the OTC rate depends endogenously on market tightness.[14]

At any round n, the bargained amount r_n^f equals the cash rate plus a weighted average of the expected continuation values of remaining unmatched after round n is over, $\chi_n^\pm$. The expected cost of an unmatched deficit in round n, χ_n^-, is in turn given by the cost of closing the deficit in the next round, $r_{n+1}^f - r^m$, multiplied by the probability of matching ψ_{n+1}^-, plus the expected cost of continuing unmatched by the end or round $n+1$, χ_{n+1}^-, multiplied by the probability $1-\psi_{n+1}^-$. An analogous recursion applies to χ_n^+ for surplus investors.

This recursive structure shows that the solution depends only on the round index n, not on the identity of the meeting counterparties. As $\Delta \to 0$, the trade size vanishes, so the impact of the transaction on each investor's equity becomes negligible, rendering the equilibrium rate independent of the traders' individual equity or surplus magnitudes.

We prove this result formally in Appendix A.1, but we outline the main steps here. When a trade of size Δ occurs at rate r_n, the change in utility for investor j with position s^j satisfies

$$\lim_{\Delta\to 0^+} \frac{V(\mathcal{E}^j + \text{sign}(s^j)(r_n - r^m)\Delta) - V(\mathcal{E}^j + \text{sign}(s^j)\chi_n^{\text{sign}(s^j)}\Delta)}{\Delta} = V'(\mathcal{E}^j)\text{sign}(s^j)(r_n - r^m - \chi_n^{\text{sign}(s^j)}).$$

[14] In the last round, the OTC rate is pinned down entirely by r^m and r^w and bargaining powers and is independent of market tightness. Intuitively, given that there is no possibility of future matching, the OTC rate solves a static problem.

Hence, the Nash bargaining solution can be expressed as[15]

$$V'(\mathcal{E}^j)^\eta V'(\mathcal{E}^k)^{1-\eta} \max_r \left[\chi_n^- - (r - r^m)\right]^\eta \left[(r - r^m) - \chi_n^+\right]^{1-\eta}.$$

Because the derivatives with respect to equity $V'(\mathcal{E}^j)^\eta V'(\mathcal{E}^k)^{1-\eta}$ are independent of r, these can be factored out of the maximization. As a result, the equilibrium rate is independent of counterparties' identities and depends only on the round index n.

Letting the total surplus from Nash bargaining be $\Sigma_n \equiv \chi_n^- - \chi_n^+$, we have the conventional Nash rule implies

$$\lim_{\Delta\to 0} \mathcal{S}_n^-(\Delta) = \chi_n^- - \left(r_n^f - r^m\right) = \eta\Sigma_n, \quad \lim_{\Delta\to 0} \mathcal{S}_n^+(\Delta) = \left(r_n^f - r^m\right) - \chi_n^+ = (1-\eta)\,\Sigma_n.$$

Equivalently,

$$r_n^f = r^m + (1-\eta)\chi_n^- + \eta\chi_n^+,$$

as stated in Proposition 1.

This formulation greatly simplifies the analysis: as $\Delta \to 0$, the OTC rate depends only on the round index n, and all heterogeneity across trading pairs vanishes. The equilibrium sequence $\{r_n^f\}_{n=1}^N$ can thus be computed from a system of linear recursions in $\{\chi_n^+, \chi_n^-\}$. We also have the following corollary that confirms that the convenience-yield function (7) delivers the exact cost or benefit of settling in the OTC market as trade size converges to zero.

Corollary 1 (Convenience-yield function and coefficients)**.** *The probability of closing deficit and surplus positions in the OTC market are given by:*

$$\Psi^- = 1 - \prod_{n=1}^N \left(1 - \psi_n^-\right) \text{ and } \Psi^+ = 1 - \prod_{n=1}^N \left(1 - \psi_n^+\right).$$

The coefficients of the convenience yield function $\chi(s)$ are the period 0 solutions of (8) and (9) which satisfy:

$$\begin{aligned}\chi^- \equiv \chi_0^- &= \Psi^-(\overline{r}^f - r^m) + (1 - \Psi^-)(r^w - r^m),\\ \chi^+ \equiv \chi_0^+ &= \Psi^+(\overline{r}^f - r^m).\end{aligned}$$

where $\overline{r}^f$ is the average of $\left\{r_n^f\right\}$ weighted by the trade volume for any investor.

Once we take the continuous-time limit, we obtain a closed-form expression for the convenience

[15] To see this, we use that the objective is $\lim_{\Delta\to 0^+} \max_r \left\{\left[S_n^-(\Delta)/\Delta\right]^\eta \left[S_n^+(\Delta)/\Delta\right]^{1-\eta}\right\}$ and use the definitions of $\mathcal{S}$ and V.

yield function as a function of the initial and final market tightness.

3.4 Continuous-time limit

The time interval between rounds is $1/N$, where recall that we normalize the total time to one. Thus, as $N \to \infty$, we can associate a round with a point in time, $\tau \in [0,1]$. Therefore, from now on, we index all equilibrium variables by τ instead of by n.

Let us define

$$\gamma(\theta) \equiv G(\theta, 1), \quad \lambda \equiv \lim_{N\to\infty} \lambda_N N.$$

By constant returns to scale in the matching function, we have that

$$\psi^+(\theta) = \lambda\gamma(\theta) \quad \psi^-(\theta) = \lambda\gamma\left(\theta^{-1}\right), \tag{10}$$

where $\psi^+(\theta) = \theta\psi^-(\theta)$. Notice that with some abuse of notation we have transformed matching probabilities into matching rates (expected number of matches per unit time).[16] These matching rates determine the fraction of surplus/deficit that an investor will be able to patch in the OTC market:

$$\Psi_\tau^+ = 1 - \exp\left(-\int_\tau^1 \psi_x^+ dx\right), \quad \Psi_\tau^- = 1 - \exp\left(-\int_\tau^1 \psi_x^- dx\right). \tag{11}$$

where we refer to $\psi_\tau^+ \equiv \psi^+(\theta_\tau)$ and $\psi_\tau^- \equiv \psi^-(\theta_\tau)$ as the matching rates as a function of time τ. The following result characterizes the evolution of θ:

Proposition 2 (ODE for market tightness). *For any given θ_0, the ratio of deficit to surpluses θ_τ satisfies the following first-order ordinary homogeneous differential equation:*

$$\dot{\theta}_\tau = -\lambda\theta_\tau\left[\gamma\left(\theta_\tau^{-1}\right) - \gamma(\theta_\tau)\right], \quad \tau \in [0,1]. \tag{12}$$

As the number of rounds increases, θ is translated from a sequence into a function of time, solving the ODE (12).[17] Once we solve the ODE for tightness, we can solve for the matching intensities

[16] We have $\lim_{N\to\infty} \psi_n N = \psi(\theta)$, where recall that $\psi_n = \lambda_N G(1,\theta)$. The overall matching probabilities, thus, correspond to cumulative distribution functions of the event, matching between time τ and time 1. Their corresponding PDF are functions of the matching rates.

[17] The Picard-Lindelöf Theorem guarantees the uniqueness of the solution of (12), as long as γ is Lipschitz continuous. We note that for some matching functions, γ may fail to be Lipschitz continuous at $\theta = 0$ or ∞, but even in that case, uniqueness can be guaranteed by finding the path of θ backward in time to reach the initial condition θ_0 and using the fact that if $\theta_\tau = 0$ for any τ remains at zero, for all $\tau' > \tau$. Observe that the reciprocal tightness, θ^{-1}, satisfies the ODE:

$$\dot{\theta}^{-1} = -\lambda\theta^{-1}\left[\gamma(\theta) - \gamma\left(\theta^{-1}\right)\right] = -\lambda\gamma(\theta_\tau)(1-\theta_\tau), \tag{13}$$

ψ_τ^+ and ψ_τ^- in $\tau \in [0,1]$. Furthermore, with information regarding the terminal market tightness, $\bar{\theta} \equiv \theta_1$, we recover the overall matching probabilities by using the matching consistency condition:

$$\Psi_0^+ = \frac{\bar{\theta} - \theta_0}{\bar{\theta} - 1}, \quad \Psi_0^- = \frac{\bar{\theta} - \theta_0}{\bar{\theta} - 1}\theta_0^{-1}.$$

We have, in addition, the following corollary:

Corollary 2 (Properties of market tightness)**.** *The evolution of θ features the following properties:*

i. If $\theta_0 = 1$, then $\theta_\tau = 1$ for all τ. If $\theta_0 > 1$ ($\theta_0 < 1$), then θ_τ is increasing (decreasing) over time.

ii. If $\theta_0 = 1$, matching rates are $\psi_\tau^+ = \psi_\tau^- = \lambda$. If $\theta_0 > 1$ ($\theta_0 < 1$), then ψ_τ^+ is increasing (decreasing) and ψ_τ^- is decreasing (increasing) with time.

iii. Take two matching functions, such that $\gamma(\theta) < \tilde{\gamma}(\theta) \forall \theta$. Then, if $\theta > 1$ ($\theta < 1$), θ rises (falls) faster under $\tilde{\gamma}$.

The corollary highlights that the short side of the market becomes relatively more scarce over the rounds, which in turn induces a higher matching rate for the short side. Similarly, we can solve for the sequence of convenience yield coefficients:

Corollary 3 (Convenience-Yield Coefficients in the Continuous-Time Limit)**.** *The solution to $\{\chi_\tau^+, \chi_\tau^-\}$ is:*

$$\chi_\tau^+ = (r^w - r^m) \int_\tau^1 (1-\eta)\,\psi_y^+ \exp\left(\int_y^1 -\left((1-\eta)\,\psi_x^+ + \eta\psi_x^-\right) dx\right) dy$$

and

$$\chi_\tau^- = (r^w - r^m)\left(1 - \int_\tau^1 \eta\psi_y^- \exp\left(\int_y^1 -\left((1-\eta)\,\psi_x^+ + \eta\psi_x^-\right) dx\right) dy\right),$$

for all $\tau = [0,1]$.

Notice the coefficients of the convenience-yield function (7) are $\chi^+ = \chi_0^+$ and $\chi^- = \chi_0^-$ and the bargained OTC rate at round τ is given by $r_\tau^f = r^m + (1-\eta)\,\chi_\tau^- + \eta\chi_\tau^+$. The formula is equivalent to a setting where one side of the market extracts all the surplus with a constant probability. To see this, it is convenient to compute the surplus at time τ, which in this case is given by:

$$\Sigma_\tau \quad = (r^w - r^m)\,(1 - H_\tau^+)\,(1 - H_\tau^-)\,,$$

where $H_\tau^+ \equiv 1 - \exp\left(\int_\tau^1 -\left((1-\eta)\,\psi_x^+\right) dx\right), \quad H_\tau^- \equiv 1 - \exp\left(\int_\tau^1 -\left(\eta\psi_x^-\right) dx\right).$

where the second equality follows from symmetry and homogeneity, $\gamma(\theta) = \theta\gamma\left(\theta^{-1}\right)$. Thus, the ODE for tightness or its reciprocal is the same.

This representation reveals two implications. First, the surplus increases over time because, as time runs out, the option to trade in the future falls. Second, the surplus at time τ is proportional to the trading between the surplus in the final round, multiplied by the product of two probabilities, $(1 - H_\tau^+)$ and $(1 - H_\tau^-)$. The product $(1 - H_\tau^+)(1 - H_\tau^-)$ is analogous to the probability that neither side extracts any surplus in the future.[18]

We further simplify the formulas for χ_τ^- and χ_τ^+ as time-varying fractions of the total surplus:

$$\chi_\tau^+ = \int_\tau^1 (1-\eta)\,\psi_y^+ \Sigma_y dy, \quad \chi_\tau^- = (r^w - r^m) - \int_\tau^1 \eta \psi_y^- \Sigma_y dy. \tag{14}$$

Thus, the outside options satisfy a dynamic surplus splitting rule and are obtained as the expected probability of extracting all the surplus in a future match.

Numerical illustrations. Figure 1 presents examples of the OTC market using Leontief and Cobb-Douglas matching functions, for various initial conditions θ_0. The figure reports the movement of the OTC market rate through time, plotted together with the outside options. As the trading session ends, the outside options for the surplus (deficit) side become zero $(r^w - r^m)$. The outside options at the beginning of the trading session yield the convenience yield coefficients.

3.5 Analytical Properties

In this section, we present some properties of the OTC market and liquidity-yield coefficients.

Liquidity-Yield Coefficients. While the solution for the path of market tightness depends on the matching function and, thus, no general analytic solution may be found, the yield coefficients can be expressed solely as a function of the initial tightness, θ, and the terminal tightness $\bar{\theta}$, in addition to r^m and r^w.[19]

Proposition 3 (Convenience-yield coefficients)**.** *For any matching function, the convenience yield co-*

[18] Note that $\left(1 - \int_\tau^1 \frac{\partial}{\partial y} \exp\left(\int_y^1 -((1-\eta)\,\psi_x^+ + \eta\psi_x^-)\,dx\right)\right) dy = \exp\left(\int_\tau^1 -((1-\eta)\,\psi_x^+ + \eta\psi_x^-)\,dx\right)$ corresponding to the probability that an (homogeneous) Poisson event is *not* realized between τ and 1. Indeed, H_τ^+ and H_τ^- are the probabilities of two independent Poisson events occurring with respective intensities $(1-\eta)\,\psi_x^+$ and $\eta\psi_x^-$. Thus, H_τ^+ and H_τ^- correspond to a compound process in which borrowers and lenders extract all the terminal surplus $(r^w - r^m)$ with probability $1-\eta$ and η, respectively, once they match.

[19] Note, however, that $\bar{\theta}$ depends on the actual matching function.

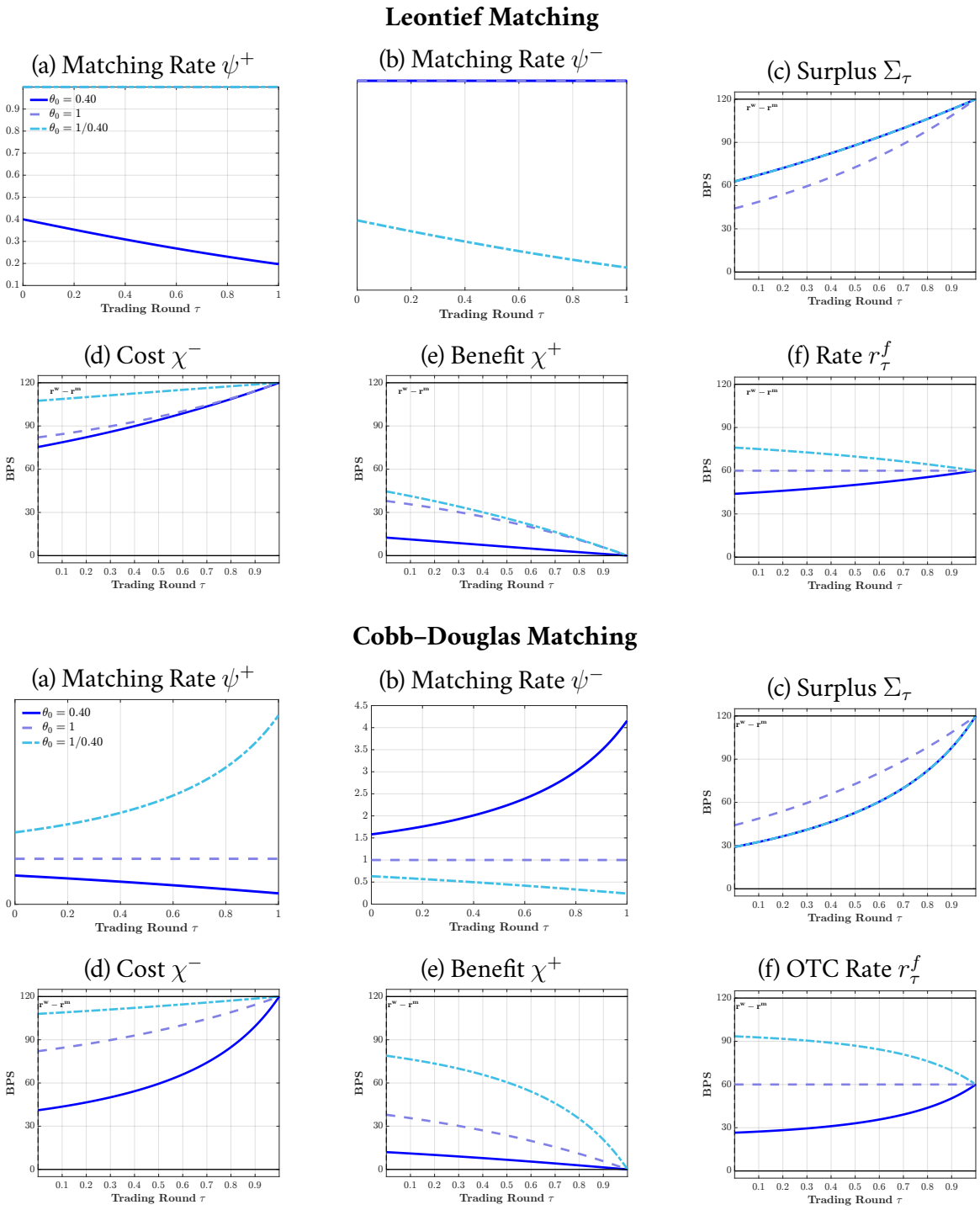


Figure 1: Trading Across Rounds under Different Matching Technologies.

Note: Parameter values are $\eta = 0.5$, $\lambda = 1.2$, $r^w - r^m = 120$bps

efficients are:

$$\chi_\tau^+ = (r^w - r^m)\left(\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{1-\eta}}{\bar{\theta} - 1}\right), \quad \chi_\tau^- = (r^w - r^m)\left(\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{-\eta}}{\bar{\theta} - 1}\right).$$

Moreover, the average OTC rate is given by:

$$\overline{r}^f = \phi(\theta_0)\, r^m + (1 - \phi(\theta_0))r^w, \text{ where } \phi(\theta_0) = \frac{\left(\bar{\theta}/\theta_0\right)^\eta - 1}{\bar{\theta}/\theta_0 - 1} \in [0, 1].$$

The average OTC rate is a weighted average of the cash rate and penalty rate. The term $\phi(\theta_0)$ acts as an endogenous bargaining rate: unlike static bargaining, this bargaining rate depends on the evolution of outside options, as in Afonso and Lagos (2015b).

As a corollary, we have that the time pattern of variables in Figure 1, corresponding to Leontief and Cobb-Douglas matching functions, holds for all matching functions:

Corollary 4 (Monotonicity Within Trading Session). *For all $\tau \in (0, 1)$, we have $\frac{d\chi_\tau^+}{d\tau} < 0$ and $\frac{d\chi_\tau^-}{d\tau} > 0$. Moreover, sign $\left(\frac{dr_\tau^f}{d\tau}\right) = -$sign$(\theta_0 - 1)$.*

Notice that when the OTC market is balanced, $\theta_0 = 1$, the market remains balanced throughout and, thus, matching rates are equalized: $\psi_\tau^+ = \psi_\tau^- = (1 - \eta)\psi_\tau^+ + \eta\psi_\tau^- = \lambda$. It follows that $\overline{r}^f = \eta r^m + (1 - \eta)r^w$.

Time Dilation. The passage of time is equivalent to a reduction in efficiency, a property which we refer to as time dilation. Let $\theta(\tau, \theta_0, \lambda)$ denote the value of market tightness θ at time τ given an initial condition θ_0 and efficiency λ, which is the solution of (12) as a function of time, initial condition and parameters. We have that:

Corollary 5 (Time Dilation). *Fix $\tau, \tau' \in [0, 1]$ such that $\tau' > \tau$. Then,*

$$\theta(\tau', \theta_0, \lambda) = \theta\left(\frac{\tau' - \tau}{1 - \tau}, \theta(\tau, \theta_0, \lambda), \lambda(1 - \tau)\right).$$

The same property holds for $\{\gamma^+, \gamma^-, \chi^+, \chi^-, r^f\}$.

Time dilation allows us to obtain the value of market tightness at an instant τ', by computing first the value of market tightness at a prior instant τ. In particular, we can obtain the value at τ' by (i) solving the equilibrium renormalizing time by the remaining time $(1 - \tau)$, (ii) scaling efficiency by the remaining time, and (iii) setting the initial condition to θ_τ. This property is useful because it

tells us that any property (e.g., monotonicity, concavity, etc.) of the equilibrium functions at some moment in time generalizes to all times. It also reveals the recursive nature of the market and shows that the normalization of time to the interval $[0, 1]$ is inconsequential.

Bargaining Power. Next, we characterize how bargaining power affects outcomes.

Corollary 6 (Role of Bargaining Power). *The equilibrium objects* $\{\chi^+, \chi^-, \bar{r}^f\}$ *are decreasing in* η. *In addition, at the extremes, we have that*

i. $\eta = 1$

$$\bar{r}^f = r^m, \quad \chi^+ = 0, \quad \chi^- = (1 - \Psi^-)(r^w - r^m),$$

ii. $\eta = 0$

$$\bar{r}^f = r^w, \quad \chi^+ = \Psi^+(r^w - r^m), \quad \chi^- = r^w - r^m.$$

An increase in borrowers' bargaining power, η, leads to a lower OTC rate at every round. Thus, convenience yields are decreasing. In addition, as borrowers extract all the matching surplus, rates go to the cash rate, and the average cost of deficits is the penalty rate multiplied by the probability of matching throughout all rounds, $\chi^- = (1 - \Psi^-)(r^w - r^m)$. By symmetry, an analogous result holds in the opposite extreme.

Matching efficiency limits. We also derive the limiting properties as the market efficiency is taken to its extreme values.

Corollary 7 (Matching efficiency limits). *The OTC market equilibrium satisfies:*

Walrasian: *As* $\lambda \to \infty$, *the OTC market converges to its Walrasian market:*

i) If $\theta_0 = 1$*:*

$$\Psi^+ = \Psi^- = 1, \quad \chi^+ = (r^w - r^m)(1 - \eta), \quad \chi^- = (r^w - r^m)(1 - \eta),$$
$$\bar{r}^f = r^m + (1 - \eta)(r^w - r^m).$$

ii) If $\theta_0 > 1$*:*

$$\Psi^+ = 1, \Psi^- = \theta_0^{-1}, \quad \chi^+ = \chi^- = r^w - r^m, \quad \bar{r}^f = r^w.$$

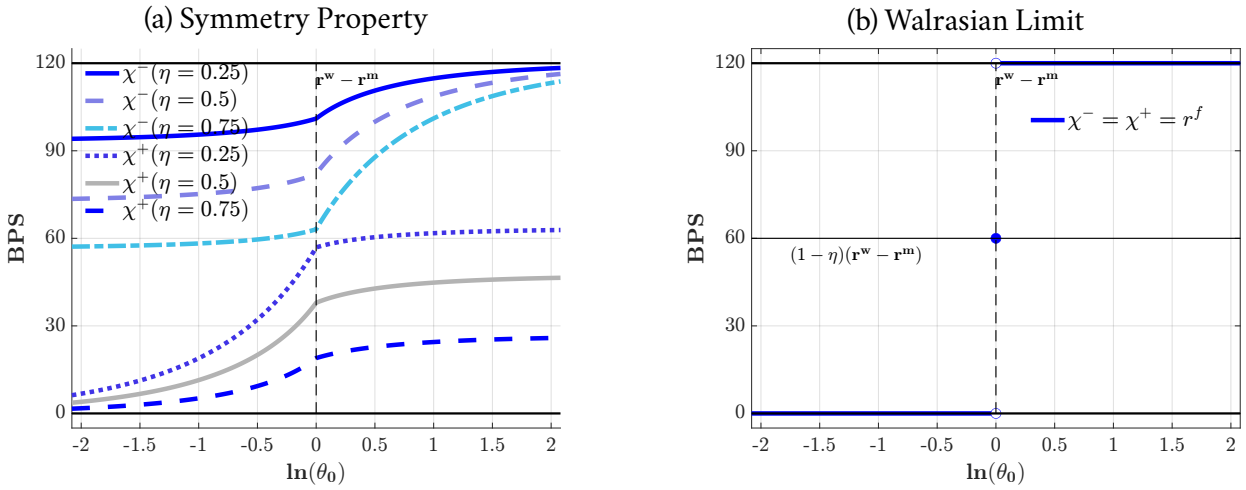


Figure 2: Symmetry and Walrasian Limit Properties

Note: Convenience yield coefficients as functions of θ_0. Trading Across Rounds. Parameter values are $\lambda = 0.8$, $r^w - r^m = 120$bps in panel (a) and $\eta = 0.5$, $r^w - r^m = 120$ bps in panel (b).

iii) If $\theta_0 < 1$:

$$\Psi^+ = \theta_0, \quad \Psi^- = 1, \quad \chi^+ = \chi^- = 0, \quad \overline{r}^f = r^m.$$

Autarkic: *As $\lambda \to 0$, the OTC market converges to a static bargaining:*

$$\chi^+ = 0, \quad \chi^- = r^w - r^m, \quad \overline{r}^f = r^m + (1-\eta)\left(r^w - r^m\right).$$

Corollary 7 shows that as efficiency increases, the OTC market rate approaches a Walrasian limit.[20] In the Walrasian limit, if the market features an aggregate cash (scarcity of funds) deficit, $\theta_0 > 1$, the average OTC market rate converges to the borrowers' outside option, r^w. In the opposite case, as $\theta_0 < 1$, the rate converges to r^m. Likewise, the liquidity yields converge to the terminal trading surplus (in the case of an aggregate cash deficit) and zero (otherwise). Panel (b) of Figure 2 plots the rates and convenience yields at the Walrasian limit.[21] On the other hand, when matching efficiency approaches zero, rates and yields converge to those of the static bargaining, and trade volume vanishes.

Market Tightness. Another property of interest regards how the convenience yield function varies with market tightness. We have the following corollary:

[20]The workhorse Poole (1968) model is Walrasian but it generates the downward sloping relation between liquidity and the interbank market rate by assuming that the interbank market closes before withdrawal shocks are realized.

[21]In the knife-edge case where $\theta = 1$, the average rate is an average of outside options weighted by the bargaining power. In either case, the trading probability equals 1 for the shortest side of the market.

Corollary 8 (Monotonicity in market tightness). *In any OTC market equilibrium,* $\left\{\chi^+, \chi^-, \bar{r}^f\right\}$ *are increasing in* θ.

As deficits increase, lenders charge more for their funds and can match with greater probability. Borrowers pay higher interest rates on their OTC borrowings and are more likely to borrow at penalty rates. Hence, $\chi^+, \chi^-, \bar{r}^f$ are increasing in θ.

Next, we investigate the limit as market tightness approaches its extrema. By symmetry, it is sufficient to discuss the limit $\theta_0 \to 0$.[22] Understanding this limit is important because it tells how the market behaves as the *aggregate* cash deficit vanishes. This limit highlights some subtle features of the model. One might initially expect that as one side of the market vanishes—say, as $\theta_0 \to 0$—the short side (i.e., borrowers) would capture the full surplus, leading rates and yield coefficients to approach zero. However, the actual limiting behavior is more nuanced, and this intuition holds only under specific conditions.

Whether the shortest side of the market extracts all surplus depends on whether the intensive form of the matching function $\gamma(\cdot)$ is bounded above. Let $\bar{\gamma} \equiv \lim_{\theta \to 0} \gamma(\theta^{-1})$. Clearly, $\bar{\gamma}$ is bounded for some matching functions, e.g., for the harmonic mean, $G(a,b) = \left(\frac{1}{2}a^{-1} + \frac{1}{2}b^{-1}\right)^{-1}$ but not for others, e.g., the Cobb-Douglas, $G(a,b) = a^{1/2}b^{1/2}$. This bound is critical in determining the decay rate of θ as θ vanishes:

$$\frac{\dot{\theta}}{\theta} = -\lambda\bar{\gamma} \quad \text{as} \quad \theta \to 0.$$

When the decay rate of market tightness is finite, market tightness behaves as an exponentially decaying function for θ close to zero. That is, the deficit side of the market approaches zero, but remains strictly positive for all τ as $\theta \to 0$. By contrast, if $\bar{\gamma}$ is unbounded, the decay rate explodes, so the market tightness approaches zero extremely fast as $\theta \to 0$, leading to a singularity point in the ODE where θ actually reaches zero. Thus, θ_τ in that case may approach zero in finite time. This rate of decay, in turn, governs the behavior of the OTC rates at the extrema.

Proposition 4 (Tightness Limits). *The convenience-yield function and the average OTC rate have the following limiting behavior:*

i. $\theta_0 \to 0$*:*

$$\Psi^+ = 0, \qquad \Psi^- = 1 - e^{-\lambda\bar{\gamma}},$$
$$\chi^+ = 0, \qquad \chi^- = (r^w - r^m)e^{-\lambda\bar{\gamma}\eta},$$
$$\bar{r}^f = r^m + (r^w - r^m) \cdot \frac{e^{(1-\eta)\lambda\bar{\gamma}} - 1}{e^{\lambda\bar{\gamma}} - 1}.$$

[22] See Appendix C.1.

ii. $\theta_0 \to \infty$*:*

$$\Psi^+ = 1 - e^{-\lambda\bar{\gamma}}, \qquad \Psi^- = 0,$$
$$\chi^+ = (r^w - r^m)\left(1 - e^{-(1-\eta)\lambda\bar{\gamma}}\right), \qquad \chi^- = r^w - r^m,$$
$$\bar{r}^f = (r^w - r^m) \cdot \frac{1 - e^{-(1-\eta)\lambda\bar{\gamma}}}{1 - e^{-\lambda\bar{\gamma}}}.$$

As tightness approaches zero, the lenders' matching probability approaches zero. Thus, $\chi^+ \to 0$. However, for the deficit side, the short side of the market, whether the overall probability of matching in the remaining rounds approaches one or remains less than one, depends on the asymptotic decay rate, $\bar{\gamma}$. If this decay rate is finite, although the deficit side is negligible relative to the surplus side, the matching probability converges, $1 - e^{-\lambda\bar{\gamma}}$, a number strictly less than one. As a result χ^- does not vanish. If the asymptotic decay rate is unbounded, the matching probability does converge to one and, in that case, χ^- will vanish. By symmetry, the opposite occurs in the limit as $\theta_0 \to \infty$. These properties reflect on the OTC market rate.

Notice that if the asymptotic decay rate is infinite, as market tightness vanishes, the average OTC rate approaches zero, which is akin to giving all the bargaining power to the deficit side. Instead, if the asymptotic decay rate is finite, a trader with surplus is able to extract some surplus in the improbable event that it gets to match. This occurs because the deficit side has a positive probability of not finding matches when $\bar{\gamma}$ is bounded. By symmetry, the opposite occurs as the market tightness explodes.

3.6 Closed-form solutions with CES Matching Functions

The properties above apply to all matching functions that satisfy Assumption 1. A special case is the constant-elasticity of substitution (CES) class (also known as generalized means):[23]

$$G(a, b; p) = \left(\frac{1}{2}a^p + \frac{1}{2}b^p\right)^{\frac{1}{p}}.$$

If we additionally impose concavity in the arguments of the matching function, we have that $p \leq 0$.[24] As $p \to 0$, the matching function converges to a Cobb-Douglas (geometric mean), $G(a, b; p) =$

[23] Aczél (1989, Ch. 17) shows that if, in addition to our assumptions, we impose *bisymmetry* ($G(G(a,b), G(c,d)) = G(G(a,c), G(b,d))$), then there exists a strictly increasing function f such that $G(a,b) = f^{-1}\left(\frac{f(a)+f(b)}{2}\right)$. Such functions are called *quasi–arithmetic means* or Kolmogorov means. Moreover, if G is also homogeneous of degree one, then it must belong to the CES family (see Burai, Kiss and Szokol, 2021).

[24] The coefficient relates to the elasticity of substitution $\rho \geq 0$ in production via $p = 1/(1-\rho)$. For $p \leq 0$, we have $2^{1/p} G(a, b; p) \leq \min(a, b)$, so $\lambda_N < 2^{1/p}$ guarantees weak exhaustion in the finite rounds case.

$a^{1/2}b^{1/2}$, and for $p = -\infty$, it converges to the Leontief matching function, $\min\{a, b\}$.[25]

Figure 3 compares CES matching functions. Panel (a) plots $\dot{\theta}/\theta$ as a function of θ, showing that—except for the Cobb-Douglas case—the growth rate converges, indicating finite decay rates. Panel (b) traces the time path θ_τ from a common initial condition θ_0. Only under Cobb-Douglas does θ_τ reach zero; in all other cases, tightness decays toward zero but never vanishes in finite time.[26] The plots illustrate Proposition 2: tightness decays faster for higher values of p.

Panels (c) and (d) show convenience yields as a function of $\log(\theta_0)$. Only for the Cobb-Douglas case the yield coefficients vanish (i.e., $r^w - r^m = 0$) for sufficiently small θ_0. This underscores that Cobb-Douglas is a knife-edge case within the CES family ($p \leq 0$), uniquely featuring infinite decay. Hence, it is the only specification where OTC rates and liquidity premia can be fully extinguished in finite time, provided that the market is sufficiently unbalanced. In that sense, the Cobb-Douglas matching function can deliver full cash-satiation: all deficits may be matched with surpluses, leading the OTC rate to r^m. For other matching functions, this is only possible if there are no investors in deficit and, consequently, there is no motive for trade.

Analytical Cases: Cobb-Douglas and Leontief. We are not aware of general solutions to the ODE for tightness, (12). However, when the matching function is either Cobb-Douglas or Leontief, the evolution of market tightness can be characterized analytically.[27] Table 1 presents the corresponding solutions.

The Cobb-Douglas and Leontief cases display qualitatively distinct trading dynamics. Under Leontief matching, the short side's trading probabilities are exponentially distributed, independently of the initial tightness. Thus, θ_τ never vanishes entirely throughout the trading session. Instead, θ_τ follows a logistic path. In the Cobb-Douglas case, the short side of the market may vanish before the trading time is over, if the stop time T in Table 1 is $T < 1$; evaluating the formula for θ_τ at the stopping time, $\tau = T$, yields zero or infinity, which is consistent with the end of trade.[28] In particular, under Cobb-Douglas matching, market tightness features the following Walrasian thresholds:

$$\theta^+(\lambda) \equiv \left(\frac{e^\lambda - 1}{e^\lambda + 1}\right)^2 \quad \text{and} \quad \theta^-(\lambda) \equiv \frac{1}{\theta^+(\lambda)} = \left(\frac{e^\lambda + 1}{e^\lambda - 1}\right)^2.$$

If the initial tightness is sufficiently unbalanced, $\theta_0 \notin [\theta^+(\lambda), \theta^-(\lambda)]$, trade will stop before the final

[25] Other common special cases are, $p = -1$, so that the matching function becomes the harmonic mean: $G(a, b; p) = 2\left(\frac{1}{a^{-1}+b^{-1}}\right)$. It is also known that for $q > p$, $G(a, b; p) < G(a, b; q)$.

[26] Indeed, the solution to θ_τ may reach zero in some finite time $\tau < 1$ if and only if $p = 0$ (the matching function is Cobb-Douglas).

[27] A closed form can also be found for the Harmonic matching function.

[28] The solution to θ_τ is associated with the hyperbolic tangent function, thereby featuring singularities.

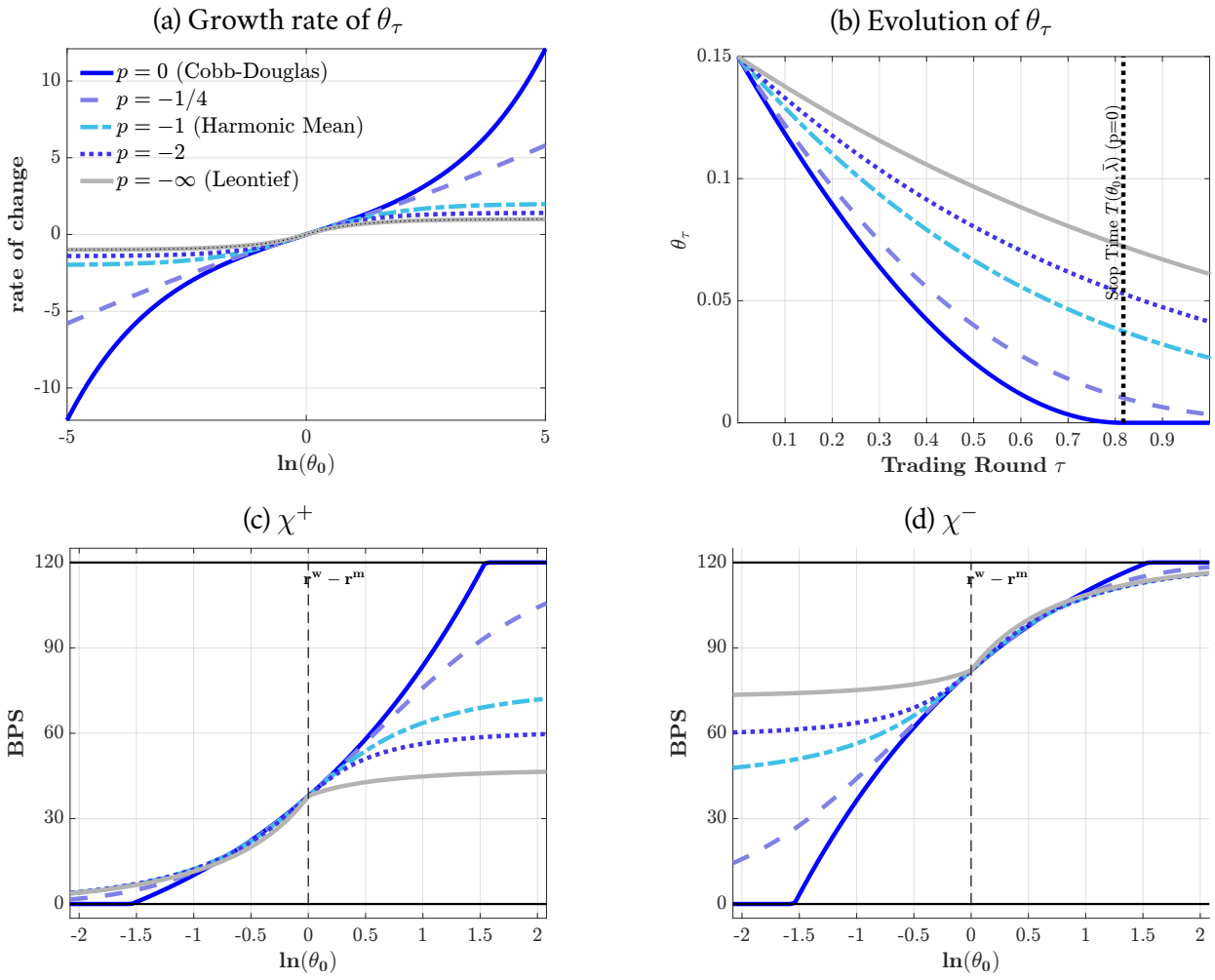


Figure 3: Comparison across CES matching functions.

Note: Parameter values are $\eta = 0.5$, $\lambda = 1.2$, $r^w - r^m = 120$bps in all cases.

trading round, and OTC rates will coincide with the Walrasian limit. Equivalently, the Cobb-Douglas case exhibits *a Walrasian efficiency threshold*:

$$\lambda^\star(\theta_0) \equiv \log\left(\left|\frac{1+\sqrt{\theta_0}}{1-\sqrt{\theta_0}}\right|\right),$$

such that trade vanishes in finite time if $\lambda > \lambda^\star(\theta_0)$.

The behavior of the yield coefficients and the average rate as a function of $\log(\theta_0)$ is depicted in Figure 4. Panels (a) and (b) respectively depict the outcomes as functions of the initial condition for the Leontief and Cobb-Douglas cases. Both figures have sigmoid-like patterns, but the differences regarding concavity and the limits as θ moves to its extreme are clear. In the Cobb-Douglas case, as market tightness becomes sufficiently unbalanced, falling outside of the interval of the Walrasian

limits, the short side of the market trades is fully matched. In such cases, the outcome coincides with the Walrasian limit, and the OTC rate corresponds to the outside option of the short side. That is, if investors in surplus exceed those in deficit, the OTC rate equals r^m. Conversely, if investors in deficit exceed those in surplus, the OTC rate equals r^w.

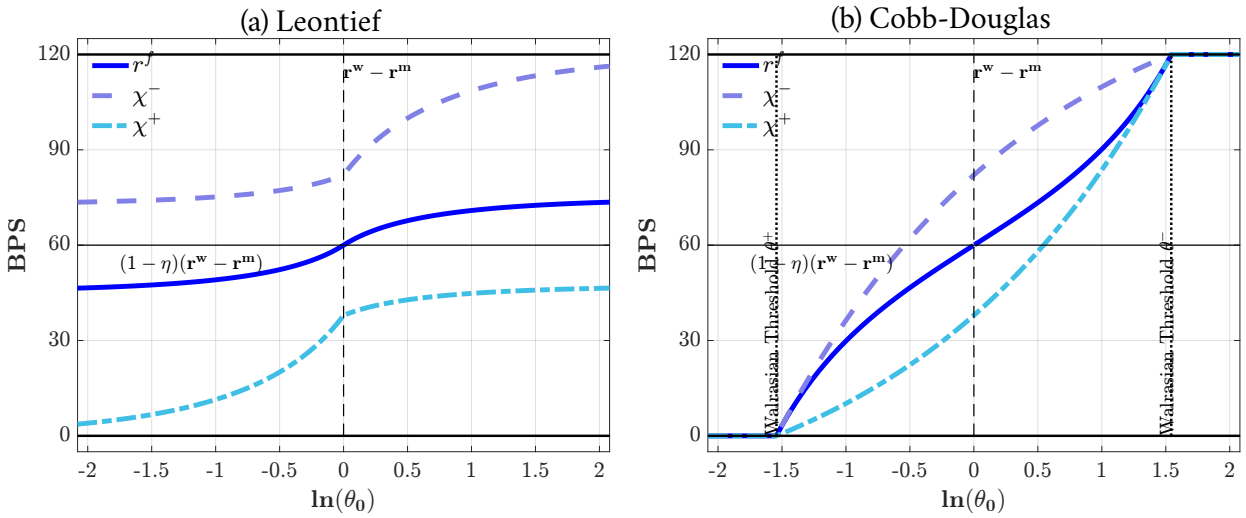


Figure 4: Analytic solutions: OTC rates and yield coefficients for Leontief and Cobb-Douglas Matching Functions.

Note: The OTC rate and convenience yield coefficients are plotted as functions of the initial market tightness, θ_0. Both panels are calibrated using $\eta = 0.5$, $\lambda = 1.2$, $r^w - r^m = 120$bps.

Table 1: Analytical Formulas:

Matching	Cobb–Douglas ($p = 0$)	Leontief ($p = -\infty$)
$\boldsymbol{\theta}(\tau),\ \tau \in [0, T]$	$\left(\dfrac{(1+\sqrt{\theta_0})\, e^{-\lambda\tau} - (1-\sqrt{\theta_0})}{(1+\sqrt{\theta_0})\, e^{-\lambda\tau} + (1-\sqrt{\theta_0})}\right)^2$	$\begin{cases} 1 + (\theta_0 - 1)e^{\lambda\tau}, & \theta_0 > 1, \\ \dfrac{\theta_0}{\theta_0 + (1-\theta_0)e^{\lambda\tau}}, & \theta_0 < 1 \end{cases}$
$\boldsymbol{T}$	$\min\left\{\dfrac{1}{\lambda}\log\left(\left\lvert\dfrac{1+\sqrt{\theta_0}}{1-\sqrt{\theta_0}}\right\rvert\right),\ 1\right\}$	∞
$\boldsymbol{\Psi^-}$	$1 - e^{-\lambda T}\left(\dfrac{(1+\sqrt{\theta_0}) - (1-\sqrt{\theta_0})e^{\lambda T}}{2\sqrt{\theta_0}}\right)^2$	$\min\{1,\ \theta_0^{-1}\}\,(1 - e^{-\lambda})$

Note: Closed-form solutions for $\theta(\tau)$, stopping time T, and deficit-side matching probability Ψ^-. The surplus-side probability satisfies $\Psi^+ = \theta_0\, \Psi^-$ for all matching functions.

4. Applications

We now study four applications of the framework. In Section 4.1, we introduce the microfounded convenience yield function into a portfolio problem and show how this affects liquidity premia. In Section 4.2, we discuss the identification of the parameters of the OTC market. In Section 4.3 we present a normative analysis of asset portfolios. In Section **??** we use the framework to connect with monetary policy implementation.

4.1 Portfolio Choices and Convenience Yields

We now study the choice of optimal portfolios. We maintain the assumption that all returns are exogenous, except for the $\overline{R}^f$ that is determined endogenously in the OTC market. Investor preferences are represented by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t),$$

where $\beta < 1$ is the time discount factor, and $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ is the utility function over the consumption good with $\gamma \geq 0$. As described in Section 2, each period, investors start with an initial equity e, which depends on previous portfolio choices and the realization of returns, and choose their portfolio decisions subject to the budget constraint (1). Below, we present the problem of the investor in recursive form. We use X to denote an aggregate state.

$$V_t(e, X) = \max_{\left\{c, \left\{\tilde{a}^i_{t+1}\right\}_{i\in\mathbb{I}}, \tilde{m}_{t+1}\right\}} u(c) + \beta \mathbb{E}\left[V_{t+1}(e_{t+1}, X')\right],$$

subject to (1), (3), and (6).

In the investor's problem, $\mathbb{E}$ denotes the expectation operator with respect to the idiosyncratic liquidity shock in the settlement stage in the current period and the next period's realization of returns.

This problem satisfies aggregation: because preferences are homothetic and equity returns scale linearly with equity, it admits a representative agent representation. Moreover, the problem admits portfolio separation: consumption (or dividends) can be chosen independently of portfolio weights, which themselves depend only on returns and settlement risks, not on total equity.[29]

Because the problem satisfies aggregation and portfolio separation, we can solve the problem of an investor with one unit of equity—after choosing consumption. This representative investor chooses a assets which amount to portfolio weights since remaining equity is a unit, $\{a^i\}$ and cash

[29] See Proposition 3 of Bianchi and Bigio (2022) for a full characterization.

m to maximize the expected utility of returns, incorporating both standard asset returns and convenience yields arising from settlement frictions:[30]

$$\max_{m,\left\{a^i\right\}_{i\in\mathbb{I}}} \left(\mathbb{E}\left[\sum_{i\in\mathbb{I}} R^i_{t+1}\left(X'\right) a^i + R^m_{t+1}(X')m + \chi_{t+1}\left(s\left(\left\{a^i\right\}_{i\in\mathbb{I}}, m, \left\{\omega^i_{t+1}\right\}_{i\in\mathbb{I}} \right); \theta_t \right) \right]^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{15}$$

subject to:

$$\sum_{i\in\mathbb{I}} a^i + m = 1,$$

where we make explicit the dependence of the convenience yield on the market tightness, with a slight abuse of notation.[31]

Figure 5 illustrates the forces underlying the portfolio problem. The kinked line represents $\chi\left(s;\theta\right)$, which maps settlement positions s into payoffs. When $s < 0$ (settlement deficit), investors face a borrowing cost χ^-; when $s > 0$ (surplus), they earn lending returns with slope χ^+. The asymmetry in the yield coefficients reflects higher penalty rates for emergency borrowing than returns on surplus lending. The panel shows two settlement distributions arising from different portfolio choices. The high-risk portfolio (red, dashed) generates a wider distribution of settlement needs, with significant probability mass in deficit regions. The low-risk portfolio (blue, solid) concentrates settlement positions at higher values. Due to the kink at zero, the high-settlement risk portfolio incurs expected losses even when distributions have zero mean. This captures a key result: assets that generate volatile settlement needs command convenience yields to compensate for these expected losses, even for risk-neutral investors.

[30]The weights are defined as portfolio holdings relative to a unit of equity. With abuse of notation, we denote the weights with the same notation as the holdings.

[31]Importantly, bounded solutions to the portfolio require that $R^i(X') < R^w$ for some X'—otherwise the investor could set $a^i \to \infty$ and borrow at the penalty rate making infinity profits.

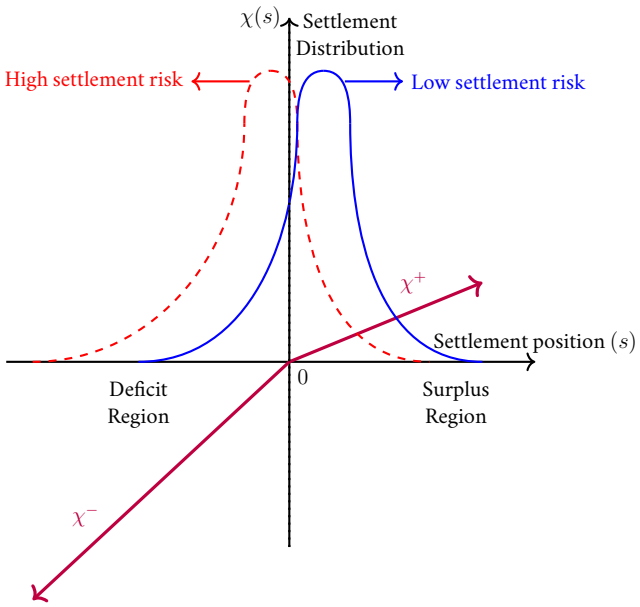


Figure 5: Convenience yield function $\chi(s)$ and settlement risk.

Convenience yields. Let us denote $\chi_s = \frac{\partial \chi}{\partial s}$, which is equal to χ^+ if $s > 0$ and χ^- if $s < 0$. Taking first-order conditions in the portfolio problem (15), and assuming strictly interior portfolio positions, we show in Appendix C.2 that this yields

$$\underbrace{\mathbb{E}_X\left[R^i\right] - R^m}_{\text{premium}} = \underbrace{\mathbb{E}_{X,\omega}\left[\chi_s \frac{\partial s}{\partial m} - \chi_s \frac{\partial s}{\partial a^i}\right]}_{\text{first-order liquidity premium}} - \underbrace{\frac{\mathbb{COV}_{X,\omega}\left[(R^e)^{-\gamma}, \chi_s \frac{\partial s}{\partial m} - \chi_s \frac{\partial s}{\partial a^i}\right]}{\mathbb{E}_{X,\omega}\left[(R^e)^{-\gamma}\right]}}_{\text{liquidity risk premium}}$$
$$- \underbrace{\frac{\mathbb{COV}_{X,\omega}\left[(R^e)^{-\gamma}, R^i(X)\right]}{\mathbb{E}_{X,\omega}\left[(R^e)^{-\gamma}\right]}}_{\text{conventional risk premium}}. \tag{16}$$

The optimal portfolio equalizes the marginal utility-weighted returns across assets, accounting for both price risk and liquidity risk due to settlement. The left-hand side is the difference in the expected return of asset i relative to cash. At an optimal solution, the representation in (16) shows that this asset's premium equals the sum of a convenience yield (liquidity premium) and a conventional risk premium. Thus, the equation captures a trade-off between expected return differentials against settlement risk and a conventional risk premium. The former emerges because asset a^i exposes the investor to settlement risk, whereas a larger m provides liquidity. In particular, in the case of a negative cash-flow shock, the investor obtains a higher return on cash compared with that of a less liquid asset, and vice versa. Assets that induce greater liquidity risk command a greater premium.

The liquidity premium term can be further unpacked into two terms. The first term captures how changes in the portfolio affect the expected settlement costs.[32] This term is present under risk neutrality and always positive. In turn, the second term captures the covariance term between liquidity payoffs and the discount factor: when the investor is risk-averse, states with negative returns penalize additional settlement costs.

We highlight that cash-flow risk induces determinate portfolios even among risk-neutral investors. Unlike standard portfolio problems where the riskiness of assets is given, here, by choosing their cash position, investors control the amount of risk. Given the concavity of χ, investors must be compensated with returns in order to hold portfolios that are more exposed to settlement risk. In addition, the standard risk-adjustment is insufficient to price the risk associated with an asset. Moreover, convenience yields cannot be treated as pricing factors independent of risk, because the correlation between the riskiness of the asset and the liquidity needs must be considered.

4.2 Comparative Statics and Parameter Identification

An important empirical question is how to identify the underlying OTC market parameters and settlement risk from observable data. Our analytical formulas reveal that the average OTC rate $\bar{r}^f$ and the convenience yield coefficients, χ^+, χ^-, depend on three key elements: market tightness θ, matching efficiency λ, and bargaining power η. While these parameters may not be directly observable, our framework provides a way to link them to measurable outcomes such as the OTC rate dispersion, the spread between the average OTC rate and the cash rate, other convenience yields, and measures if trading volume. One particular measure of trading volume is the ratio of penalty funding volume to OTC market funding volume:

$$I(\theta) \equiv \frac{1 - \Psi^-(\theta)}{\Psi^-(\theta)}.$$

This ratio captures market efficiency in reallocating liquidity: if $I(\theta) = 0$ indicates perfect reallocation through the OTC market whereas $I(\theta) = \infty$ indicates no OTC volume at all. We call $I(\theta)$ simply the *relative volume*.

(i) *Market tightness:* Consider first the role of market tightness. Recall that, as established in Corollary 8, the convenience yield coefficients and the average OTC rate $\chi^+, \chi^-, \bar{r}^f$ are all monotonically increasing in market tightness θ. The monotonicity suggests that one can, in principle, back out θ from observed rates.

[32] Note that $\chi(s)$ is continuous and piecewise linear, with slope χ^- for $s < 0$ and χ^+ for $s > 0$; hence it is differentiable everywhere except at $s = 0$. Changing the portfolio affects, at the margin, the probability of being in deficit. However, this does not enter the optimality condition: by Leibniz's rule, the derivative at the threshold ω^* vanishes because the bank's marginal payoff is zero at that point.

We also observe that the relative funding volume is increasing in θ, given that Ψ^- decreases with θ. However, for the Leontief case, when markets have excess liquidity ($\theta < 1$), relative volume is independent of θ:

$$I(\theta_0) = \begin{cases} \frac{e^{-\lambda}}{1-e^{-\lambda}} & \theta \leq 1 \\ \\ \frac{1-\left(1-e^{-\lambda}\right)\theta^{-1}}{\left(1-e^{-\lambda}\right)\theta^{-1}} & \theta > 1. \end{cases}$$

Thus, the relative volume will not be informative about θ for matching functions close to the Leontief.

ii) *Matching efficiency:* Let us now turn to the comparative static with respect to the efficiency parameter λ. We first proceed by establishing how the efficiency parameter affects market tightness:

Corollary 9 (Comparative statics in matching efficiency)**.** *Fix an initial market tightness θ_0. Then,*

$$\operatorname{sign}\left(\frac{\partial \theta_\tau}{\partial \lambda}\right) = \operatorname{sign}(\theta_0 - 1), \forall \tau \in (0, 1]\,.$$

The corollary establishes that if we start from a market tightness $\theta_0 > 1$, then θ_τ is strictly increasing in λ at any trading round. Conversely, if $\theta_0 < 1$, it is strictly decreasing in λ. This result is intuitive: higher matching efficiency accelerates the clearance of the short side of the market, which amplifies the relative scarcity of that side and therefore increases (decreases) market tightness when $\theta_0 > 1$ ($\theta_0 < 1$). As a result, changes in matching efficiency affect all equilibrium objects solely through their impact on the path of market tightness. Indeed, this is established in the following proposition:

Proposition 5 (Comparative Statics about λ)**.** *Efficiency changes the equilibrium objects $\left\{\chi^+, \chi^-, \overline{r}^f\right\}$ as follows:*

i) *Let $\theta_0 < 1$, then $\overline{r}^f$ is decreasing in λ, χ^- is decreasing in λ, and χ^+ is non-monotonic in λ (first decreasing and then increasing).*

ii) *Let $\theta_0 > 1$, then $\overline{r}^f$ is increasing in λ, χ^+ is increasing in λ, and χ^- is non-monotonic in λ (first increasing and then decreasing).*

iii) *Let $\theta_0 = 1$, then χ^+ is increasing in λ, and χ^- is decreasing in λ.*

The comparative static reveals a subtle but important insight: convenience yields may be non-monotonic in the matching efficiency. When markets are tight ($\theta_0 > 1$), improving matching efficiency has two opposing effects: it increases the probability that deficit investors find lenders (reducing their borrowing costs), but it also strengthens lenders' bargaining position, raising equilibrium

rates. The net effect on χ^- depends on which force dominates. This non-monotonicity implies that higher convenience yields need not indicate market dysfunction but may instead reflect improved matching that benefits the scarce side of the market. Figure 6 illustrates these relationships: Panel (a) confirms our theoretical results for the Leontief case, while Panel (b) demonstrates that this pattern extends to the Cobb-Douglas matching function.

The non-monotonicity of rates and liquidity premia with respect to efficiency implies that convenience yields should not be interpreted mechanically as evidence of deteriorating market efficiency. To see this, consider the first-order liquidity premium in equation (16) for an asset that does not expose investors to settlement risk; that is, $\frac{\partial s}{\partial a^i} = 0$. In that case, the premium reduces to $\mathbb{E}_{X,\omega}[\chi_s]$. An increase in this premium could be consistent with two opposite interpretations. On the one hand, it might signal weaker OTC efficiency, leading investors to demand cash as insurance against settlement obligations. On the other hand, it could reflect greater efficiency, which enhances the option value of holding cash to exploit lending opportunities. Thus, higher observed convenience yields do not unambiguously point to deteriorating trading efficiency.

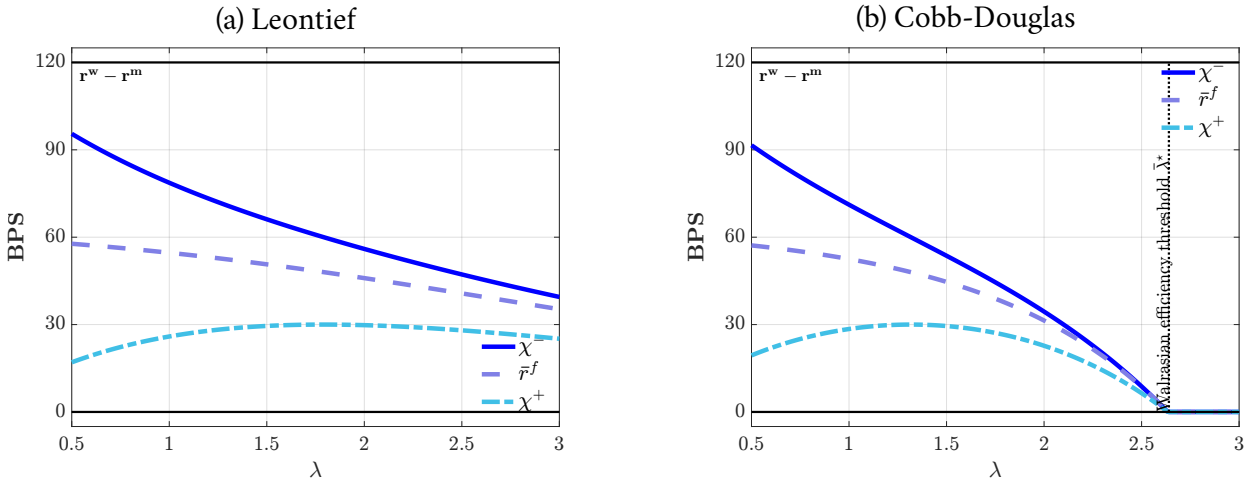


Figure 6: Effects of matching efficiency

Note: Rates and convenience yield coefficients as functions of λ for Leontief and Cobb-Douglas matching functions. The OTC rate and liquidity yield coefficients are plotted as a functions of λ. Both panels are calibrated using $\eta = 0.5$, $\theta_0 = 0.75$, $r^w - r^m = 120$bps.

iii) *OTC Rate Dispersion:* The dispersion of OTC rates within each trading day—the difference between rates at the beginning and end of the trading session—provides another observable moment for parameter identification. Recall from Proposition that the equilibrium rate varies during the trading session according to the evolution of the tightness θ_τ. Since rates change monotonically throughout the day, increasing (decreasing) over time if $\theta > 1$ ($\theta < 1$), we can measure dispersion

in rates.

$$Q \equiv \max_{\tau} r_{\tau}^{f} - \min_{\tau} r_{\tau}^{f} = \left| r_{1}^{f} - r_{0}^{f} \right|.$$

We obtain the following relationship between parameters and the dispersion:

Corollary 10 (Comparative Statics: Rate Dispersion)**.** *Rate dispersion features the following comparative statics: (i) If $\theta < 1$, then $\frac{\partial Q}{\partial \theta} < 0$; ii) If $\theta > 1$, then $\frac{\partial Q}{\partial \theta} > 0$; iii) If $\theta = 1$, then $\frac{\partial Q}{\partial \theta} = 0$; iv) $\frac{\partial Q}{\partial \lambda} \geq 0$ with equality if and only if $\theta = 1$.*

The intuition is straightforward. Rate dispersion reflects how much the market "unwinds" during the trading session. When the market is unbalanced ($\theta \neq 1$), the scarce side gets progressively better terms as trading proceeds, creating larger rate movements. The more unbalanced the market—in either direction—the greater the dispersion. Similarly, higher matching efficiency λ accelerates this unwinding process, amplifying dispersion. At $\theta = 1$, the market is balanced and rates remain constant throughout the session, yielding $Q = 0$. Figure 7 illustrates these patterns.

iv) *Relative Volumes of Market and Outside Funding:* The relative volume behaves as follows with respect to

Corollary 11 (Relative Volume and Matching Efficiency)**.** *The matching probability Ψ^{-} is increasing in λ for all $\theta_0 \neq 1$. Consequently, the relative volume $I(\theta_0) \equiv \frac{1-\Psi^{-}}{\Psi^{-}}$ is decreasing in λ for all θ_0.*

These results reveal an insight for identification: the relative volume decreases monotonically with matching efficiency, regardless of the market tightness. This contrasts sharply with convenience yields, which can be non-monotonic in λ. Combined with the observation that relative volume may be insensitive to the tightness (when $\theta < 1$ and p small), this suggests that the relative volume is convenient to pin down λ whereas tightness should be inferred from yields.

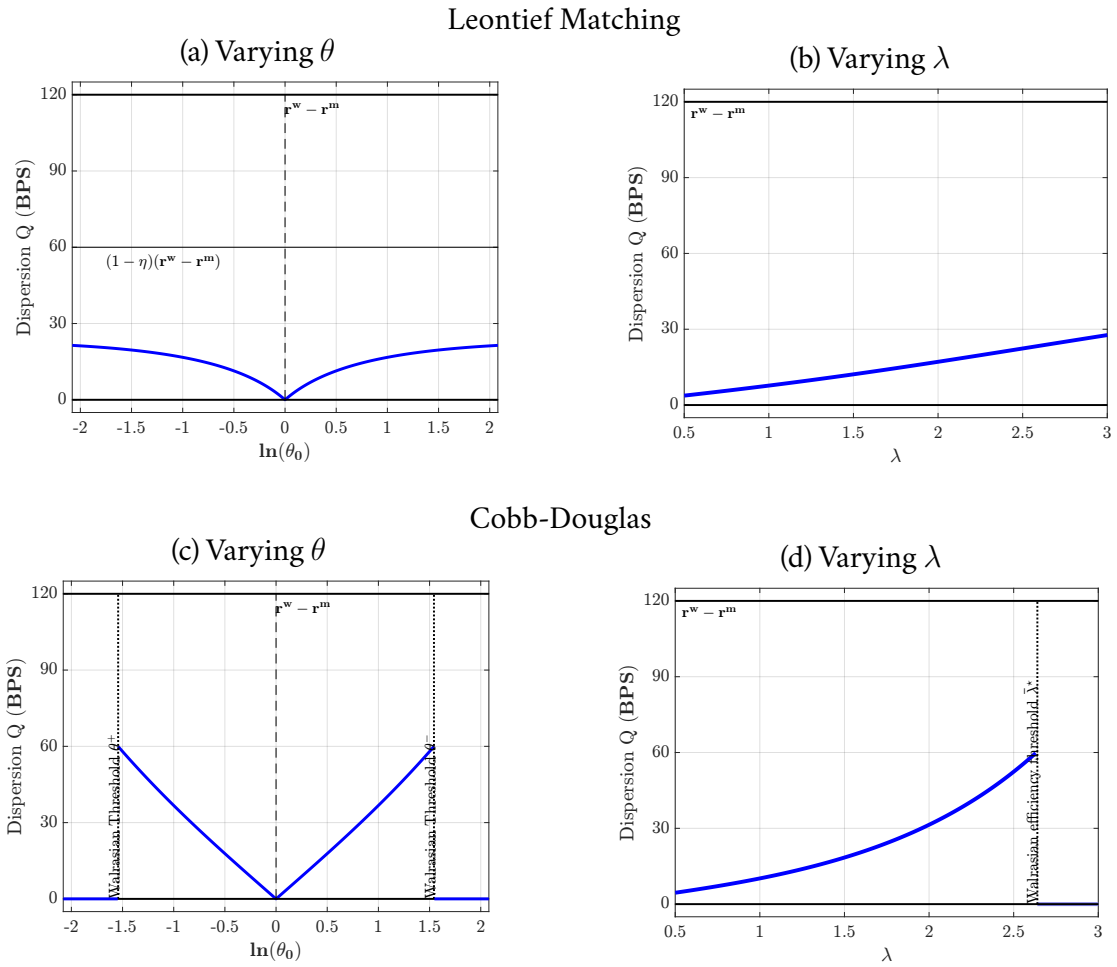


Figure 7: Dispersion of OTC Rates as a function of $\{\lambda, \theta_0\}$.

Note: The dispersion of the OTC rate is plotted as a function of λ and θ_0 for the Leontief and Cobb-Douglas matching functions. In all panels, $\eta = 0.5$, $r^w - r^m = 120$bps. In panels (a) and (c) we set $\lambda = 1.2$. In panels (b) and (d), we set $\theta_0 = 0.75$.

Summing up. Our comparative-statics results suggest a possible framework for identifying the three key parameters of the model, θ, λ, η, from observable data. This framework suggests focusing data collection on: (i) portfolio compositions across institutions, (ii) spreads between liquid and illiquid assets, (iii) the level and intraday range of OTC rates, and (iv) the relative use of emergency lending facilities, if available. We provide the sketch of the identification in the table below which can be enriched to also infer parameters of the matching function.

Algorithm 1 An Identification Approach

Given possible observable data:
- Portfolio holdings across investors, $\{m, a^i\}$, convenience yields on liquid assets: $R^a - R^m$, an average OTC rate: $\bar{r}^f$, rate dispersion: $Q = |r_1^f - r_0^f|$, relative volume

Step 1: Identify matching efficiency, λ
-Use $I(\theta)$ to infer λ

Step 2: Identify market tightness, θ**, and shock distribution,** Φ
-Given λ from Step 1:
-Use convenience yields or $\bar{r}^f$ to infer θ
-Given observed portfolios and implied θ, back out the distribution of shocks Φ

Step 3: Identify bargaining power η
-Use the historical average of $\bar{r}^f$ relative to r^m and r^w to infer η

Step 4: Validation
- Check implied $Q(\theta, \lambda)$ aligns with observed dispersion

4.3 Normative Analysis

In this section, we study the implications of the theory for liquidity regulation. We do so by comparing the decentralized portfolio choices vis-à-vis a social planner's portfolio choices. We will show that by affecting the tightness in the OTC market and the degree of congestion, the planner can improve upon the decentralized outcome.

To illustrate the externality, we consider a simple example. We specialize the model to the case where investors' portfolios are composed of a liquid asset, m, an illiquid asset, b, and deposits d. In particular, deposits are subject to withdrawal risk such that (3) is given by

$$s\left(\{b, d\}, m, \omega\right) \equiv m + \left(\frac{R^d}{R^m}\omega - \rho(1+\omega)\right) d, \tag{17}$$

here ω is distributed F with mean zero and their budget constraint $b + m = 1 + d$, where we, again, normalize equity to one.

We study the individual investor's problem next and then consider the planner's counterpart.

Investor's problem As in (15), the problem becomes the following.

$$\max_{m\geq 0, d\geq 0} \left\{ \mathbb{E}_\omega \left[R^b + R^b(d-m) + R^m m - R^d d + \chi\left(s\left(\omega, m, d\right); \theta\right)\right]^{1-\gamma} \right\}^{\frac{1}{1-\gamma}}.$$

In the example, the optimality condition (16) can be reduced to

$$\underbrace{R^b - R^m}_{\text{liquidity premium}} = \tilde{\Phi}\left(\omega^*\right)\chi^+ + \left(1-\tilde{\Phi}\left(\omega^*\right)\right)\chi^- \tag{18}$$

where

$$1-\tilde{\Phi}\left(\omega^*\right) \equiv \overbrace{\underbrace{\left(1-\Phi\left(\omega^*\right)\right)}_{\text{deficit prob}} \cdot \underbrace{\frac{\mathbb{E}_\omega\left[R^e(\omega)^{-\gamma} | \omega > \omega^*\right]}{\mathbb{E}_\omega\left[R^e(\omega)^{-\gamma}\right]}}_{\text{risk-aversion correction}}}^{\text{risk-adjusted deficit probability}}, \quad \omega^* \equiv -\frac{\rho - \frac{m}{d}}{\frac{R^d}{R^m} - \rho}.$$

Equation(18) says that the liquidity premium is a weighted average of the liquidity yield coefficients, $\tilde{\Phi}\left(\omega^*\right)\chi^+ + \left(1-\tilde{\Phi}\left(\omega^*\right)\right)\chi^-$. In the expression, $\tilde{\Phi}\left(\omega^*\right)$ is the risk-adjusted probability of falling in a cash deficit, where ω^* is the threshold shock that puts investors in deficit—recall that lower values of ω represent an outflow of funds. Thus, the liquidity premium of cash over illiquid assets is the risk-adjusted expected return of coming into the OTC market with an extra unit of cash.

In the special case of risk-neutrality, using the consistency condition of the matches, we have that (18) becomes:

$$\underbrace{R^b - R^m}_{\text{liquidity premium}} = \left(R^w - R^m\right)\left(1-\Psi^-\left(\theta\right)\right)\Phi\left(\omega^*\right).$$

Constrained-efficient allocation. We assume the planner chooses the portfolio shares (d, b, m) on behalf of investors while investors retain the choice of how much to save overall and how much to consume. The planner takes as given all portfolio returns and the structure of the OTC market, as encoded in the convenience yield function. Because investors are ex-ante identical, the problem can be expressed as follows.

$$\max_{m\geq 0, d\geq 0} \left\{ \mathbb{E}_\omega \left[R^b(d-m) + R^m m - R^d d + \chi\left(s\left(\omega, m, d\right), \theta(m,d)\right)\right]^{1-\gamma} \right\}^{\frac{1}{1-\gamma}},$$

where $s(\omega, m, d)$ is given by (17) and

$$\theta\left(m,d\right) = -\frac{\int_{-1}^{\omega^*} s\left(\omega, m, d\right)\Phi(d\omega)}{\int_{\omega^*}^{\infty} s\left(\omega, m, d\right)\Phi(d\omega)}, \quad \omega^* = \frac{\rho - \frac{m}{d}}{\frac{R^d}{R^m} - \rho}.$$

The key distinction is that the planner considers how the portfolio determines θ and how, in turn, this affects the convenience yield function χ. Notice that because all investors are identical ex ante, there are no redistribution considerations.

As shown in the Appendix C.3, the planner's first-order conditions with respect to m and b yield the following

$$\underbrace{R^b - R^m}_{\text{asset premium}} = \tilde{\Phi}\left(\omega^*\right)\chi^+ + (1-\tilde{\Phi}\left(\omega^*\right))\chi^- + \\ \frac{\partial\theta}{\partial m}\frac{\partial\chi^+(\theta)}{\partial\theta}\cdot\tilde{\Phi}(\omega^*)\cdot\mathbb{E}\left[s\cdot\frac{R^e(\omega)^{-\gamma}}{\mathbb{E}[R^e(\omega)^{-\gamma}\,|\,\omega>\omega^*]}\,\middle|\,\omega>\omega^*\right] \\ +\ \frac{\partial\theta}{\partial m}\frac{\partial\chi^-(\theta)}{\partial\theta}\cdot(1-\tilde{\Phi}(\omega^*))\cdot\mathbb{E}\left[s\cdot\frac{R^e(\omega)^{-\gamma}}{\mathbb{E}[R^e(\omega)^{-\gamma}\,|\,\omega<\omega^*]}\,\middle|\,\omega<\omega^*\right]. \quad (19)$$

As individual investors, the planner trades off the higher return on loans with the liquidity benefits of cash, considering the OTC market frictions. Unlike individual investors, however, the planner internalizes the pecuniary externality arising from how portfolio choices influence trading probabilities, as captured by the convenience yield χ.[33]

A key result is that the sign of the externality is ambiguous: it depends on the derivatives of the convenience-yield coefficients. To understand whether investors over- or under-invest in liquid assets, it is useful to consider first the limiting case with risk-neutral investors. With $\gamma \to 0$, given that $\frac{\partial\theta}{\partial m} < 0, \frac{\partial\chi^+(\theta)}{\partial\theta} > 0, \frac{\partial\chi^-(\theta)}{\partial\theta} > 0$, we have that the *planner values cash more* than individual investors if and only if

$$\frac{\partial\chi^+(\theta)}{\partial\theta}\cdot S^+ < \frac{\partial\chi^-(\theta)}{\partial\theta}\cdot S^-$$

This inequality underscores that there is an under-investment of liquid assets when the planner perceives that higher cash holdings—lower market tightness—raise the marginal return on liquid assets more in deficit than in surplus. That is, when market tightness goes up, this favors investors that are in deficit—by allowing them to borrow at a lower rate and by raising the probability of a match—relative to investors that are in surplus—as they now must lend at a lower rate and face a lower matching probability.

Suppose that the planner picks a portfolio with $\theta = 1$. Under the case with Cobb-Douglas matching, we know that $\frac{\partial\chi^+(\theta)}{\partial\theta} = \frac{\partial\chi^-(\theta)}{\partial\theta}$. If, in addition, the shock is symmetric, $F(\omega^*) = 0.5$ and $\mathbb{E}\left[-s \mid \omega < \omega^*\right] = \mathbb{E}\left[s \mid \omega \geq \omega^*\right]$.[34] In this case, it follows that there is neither over- nor under-

[33] We highlight that with multiple assets, there is no generally a Hosios (1990) condition that would render efficiency in the competitive equilibrium.

[34] To see that $\frac{\partial\chi^+(\theta)}{\partial\theta}|_{\theta=1} = \frac{\partial\chi^-(\theta)}{\partial\theta}|_{\theta=1}$. Under Cobb-Douglas, we exploit the symmetry property of derivatives.

investment of liquid assets. However, if investors were risk-averse, the risk adjustment correction would imply that by accumulating more liquid assets, the planner would effectively provide more insurance. Because individual investors do not internalize these benefits, the competitive equilibrium would feature under-accumulation of liquid assets. In addition, an allocation with a higher probability of being in deficit $\Phi(\omega^*)$ or with a more sensitive convenience yield $\frac{\partial \chi^-(\theta)}{\partial \theta}$ implies that the planner perceives a higher value from higher liquid holdings.

This externality has implications for prudential policy and liquidity regulation. Our results highlight that optimal liquidity regulation need not involve minimum liquidity requirements. In fact, depending on market conditions, the competitive equilibrium may feature over- or under-investment in liquid assets.[35]

4.4 A Taxonomy of Reserve Regimes and Interest-Rate Control

A natural application is the Feds Fund market. This section uses the results above to study when supplying reserves "abundantly" delivers precise control of the interbank rate.

Following Proposition 4, we map our framework to a taxonomy of reserve regimes, taking as given θ_0. For a full general equilibrium analysis, see Bianchi and Bigio (2022). As Figure Figure 4 illustrates, the level and responsiveness of the interbank market depends on the value of θ_0 and the matching function. Accordingly, we describe the following regimes

Definition 2 (Reserve Regimes)**.** *We say that reserves are:*

1. ***Scarce*** *if* $\theta_0 > 1$ *(deficits exceed surpluses);*
2. ***Ample*** *if* $\theta_0 < 1$ *and* $\bar{r}^f > r^m$*;*
3. ***Abundant*** *if* $\theta_0 < 1$ *and* $\bar{r}^f \to r^m$*.*

We refer to a scarce regime as one where banks' deficits exceeds surpluses. Under ample reserves, the demand curve for settlement balances is downward sloping: an increase in aggregate deficits relative to surpluses raises the OTC rate. Under abundant reserves, the demand curve is flat and the interbank market remains active, and changes in tightness have little effect on the OTC rate. Satiation is the limiting case where no investor holds a deficit position.

[35] We highlight that in our environment, the planner maximizes the investors' welfare, and the net proceeds from the trades at r^m and r^w are not rebated back to investors. Ismail and Zuniga (2025) study the problem in which the planner internalizes the revenues from discount window borrowing and rebates them to banks. Their analysis focuses on the welfare implications from an asset-demand perspective, i.e., the wedge between loan and deposit rates.

Abundant regimes. Is it possible to have a regime where the interbank market rate does not respond to market tightness, yet the market is not satiated and the interbank market rate is above r^m? Proposition 4 shows that the limiting behavior as $\theta_0 \to 0$ is governed by the asymptotic decay rate $\bar{\gamma} \equiv \lim_{\theta \to 0} \gamma(\theta^{-1})$. When $\bar{\gamma}$ is finite, e.g., in the Leontief case, abundance is achieved only asymptotically as $\theta_0 \to 0$. Moreover, the OTC spread remains strictly positive even as deficits vanish:

$$\lim_{\theta_0 \to 0} (\bar{r}^f - r^m) = (r^w - r^m) \cdot \frac{e^{(1-\eta)\lambda\bar{\gamma}} - 1}{e^{\lambda\bar{\gamma}} - 1}. \tag{20}$$

When $\bar{\gamma}$ is unbounded, as in case of Cobb-Douglas, a proper abundant regime exists. The regime is attained when m is large enough to take θ_0 below a threshold—e.g., in the Cobb-Douglas case, when $\theta_0 < \theta^+$. It follows that in an abundant regime, the short side of the market is exhausted in finite trading time. As all deficits clear before the terminal round, the OTC rate equals the rate on reserves $\bar{r}^f = r^m$ and the OTC market volume is small and, moreover, dormant for much of the trading rounds. Furthermore, the OTC market efficiently reallocates surpluses to cover all deficits and convenience yields on other assets vanish.

Figure 4 illustrates these regimes. In Panel (b), the Cobb-Douglas case exhibits a flat region where $\bar{r}^f = r^m$ for low $\theta_0 < \theta^+(\lambda)$ (the abundant regime). For tightness above that point, $\bar{r}^f$ is sensitive to the tightness (ample and scarcity regimes). In Panel (a), the Leontief case shows no flat region; rather, $\bar{r}^f$ becomes flatter and flatter as $\theta_0 \to 0$ but is never actually zero. In practice, because a proper abundant regime does not exist for matching functions with finite $\bar{\gamma}$, it only makes sense to talk about *approximate abundance* in such cases.[36] To address this, we can define an arbitrary threshold for θ_0 and designate an approximate abundance regime as any point where market tightness falls below that threshold. One practical guide is to use the threshold θ^+ from the Cobb-Douglas case as the threshold that determines the approximate abundance regime for all other CES matching functions, an ad-hoc rule we use below for comparison purposes.

Interest-Rate Control. The asymptotic decay rate $\bar{\gamma}$ also dictates whether a matching function delivers perfect interest-rate control without the need to adjust m, within an abundant regime. It is clear that if one were to hold θ_0 fixed and introduce shocks to the matching efficiency λ—or bargaining power η—, we would observe fluctuations in $\bar{r}^f - r^m$ in ample or scarcity regimes. In the case of abundant or approximately abundant regimes, the answer depends on the matching function.

For finite $\bar{\gamma}$ matching functions, a central bank does not have perfect control over the interbank

[36] This case arises under Leontief matching and, more generally, for CES matching functions with elasticity parameter $p < 0$. For all $\theta_0 > 0$, reserves are ample according to our definition. Interestingly, there may be a discontinuity at the limit $\theta_0 = 0$ in the yield coefficients. Since there is a chance deficits remained unmatched under approximate abundance but there are no deficits at the limit where $\theta_0 = 0$, there is a discontinuity in yield coefficients at the satiation point.

rate in an approximately abundant regime: Even as the interbank volume is vanishing, $\theta_0 \to 0$, the OTC rate remains sensitive to changes in λ. The intuition is that under such matching functions, deficit investors always have a positive probability of remaining unmatched. Thus, some bargaining surplus is extracted by lenders, leading to a positive OTC spread $\bar{r}^f - r^m$. However, changes in λ change that surplus extraction and, thus, the OTC rate, as is clear from equation (20).

Matching functions exhibiting an unbounded asymptotic decay rate, $\bar{\gamma} = \infty$, e.g., the Cobb-Douglas, feature the converse behavior. In this case, deficit investors cover all of their deficits with certainty for all $\theta_0 < \theta^+(\lambda)$, yielding: $\lim_{\theta_0 \to 0}(\bar{r}^f - r^m) = 0$. Crucially, the rate is invariant to λ.

Figure 6 illustrates how interest-rate control depends on the matching function. Panel (b) shows that in the Cobb-Douglas case, there is a region where the OTC rate is insensitive to changes in λ, corresponding to $\lambda > \lambda^\star$, which also defines an abundant regime in λ space. This is not the case under Leontief matching, where the OTC rate is always changes with λ.

Whether a proper abundance regime exists depends on the asymptotic decay rate of the matching function and only when a proper abundance regime exists does the central bank have perfect control over rates. Table 2 summarizes this discussion.

Table 2: Reserve Regimes under Alternative Matching Technologies

CES Matching Coefficient	**Decay Rate $\bar{\gamma}$**	**Abundant Regime**	**Rate Control**
$-\infty \leq p < 0$	Finite	Only as $\theta_0 \to 0$	Imperfect
$p = 0$ (Cobb-Douglas)	∞	$\theta_0 \in (0, \theta^+(\lambda))$	Perfect

Control vs. Sensitivity. Another property of the OTC market in the context of monetary policy implementation is a tension between how much interest-rate control the central bank has in an abundant regime and how sensitive the OTC rate is in an ample regime. We illustrate the tension by comparing CES matching functions, treating abundant or approximately abundant regimes as $\theta_0 \in (0, \theta^+(\lambda))$.

Within an abundant or approximate abundance regime, consider a change in λ. We already established that $\left|\frac{\partial(\bar{r}^f - r^m)}{\partial \lambda}\right| \geq 0$ with equality only for the Cobb-Douglas case, $p = 0$. However, shocks to λ have a lesser impact as we raise the coefficient p:

$$\lim_{\theta_0 \to 0} \frac{\partial}{\partial p} \left|\frac{\partial(\bar{r}^f - r^m)}{\partial \lambda}\right| < 0.$$

That is, the farther we move from the Cobb-Douglas case toward the Leontief case, the greater the

sensitivity of rates to shocks to efficiency.

A tension appears because the higher the coefficient of the matching function, the more sensitive is the OTC rate to changes in θ_0 in the ample reserve regime. Indeed, we have also established that $\frac{\partial(\bar{r}^f - r^m)}{\partial\theta_0} \geq 0$. However, changes in θ_0 have a greater impact as we raise the coefficient p:

$$\lim_{\theta_0 \to 0} \frac{\partial}{\partial p} \frac{\partial(\bar{r}^f - r^m)}{\partial\theta_0} > 0.$$

That is, the closer we move to the Cobb-Douglas case, the greater the sensitivity of rates to market tightness within an ample regime. Clearly, when $p = 0$ and we have Cobb-Douglas matching, the tradeoff is the starkest: Within the abundant region $\theta_0 \in (0, \theta^+)$, control is perfect—$\bar{r}^f = r^m$ regardless of shocks to λ. However, above the boundary $\theta^+(\lambda)$, the sensitivity of rates to changes in quantities is the largest.

A similar tension appears as we move toward more efficient markets. The Walrasian limit is the extreme case of this behavior: the OTC spread is zero throughout the plentiful region, but rates exhibit a discontinuous jump at the boundary of the regime.[37] Thus, in the Walrasian limit, a plentiful reserve regime coincides with the abundant regime: the central bank achieves perfect interest rate control throughout the region $\theta_0 < 1$. However, in this frictionless benchmark, the OTC spread jumps from 0 to $(1-\eta)(r^w - r^m)$ discretely when $\theta_0 = 1$:

$$\lim_{\lambda \to \infty} \frac{\partial \bar{r}^f}{\partial \theta_0} = \begin{cases} 0 & \text{if } \theta_0 < 1, \\ +\infty & \text{if } \theta_0 = 1, \\ 0 & \text{if } \theta_0 > 1. \end{cases}$$

This tension matters for policy because central banks aim to control $\bar{r}^f$. The interbank market often exhibits shocks that can relate to λ. To target the rate, they supply m to move the tightness θ_0. The greater the control in an abundant reserve regime, the less sensitive the interbank rate is to shocks and, therefore, the less the central bank has to respond to adjust to shocks. A principle emerges, however: the greater the control in an abundant reserve regime, the more sensitive are the OTC rates to changes in m in an ample reserve regime.

All in all, we obtain the following implications: (i) whether a central bank can achieve full interest rate control with an active interbank market depends on the matching technology; (ii) under finite-decay matching functions, even "approximate abundance" near satiation leaves the policy rate exposed to shocks in the OTC market; (iii) the sharper the control within an abundant regime, the

[37] Corollary 7 characterizes the Walrasian limit as efficiency approaches infinity ($\lambda \to \infty$).

more abrupt the transition between them—central banks switching from ample to abundant regimes face amplified volatility; (iv) policies or technologies that improve matching efficiency in the OTC market (e.g., centralized trading platforms, faster settlement) compress spreads and improve control within abundant regimes, but may amplify sensitivity across regimes.

5. Conclusions

We develop a tractable microfoundation for convenience yields arising from trading frictions in OTC settlement markets and show how it can be incorporated into a canonical portfolio framework. We further characterize how the convenience yield function depends on market tightness, bargaining power, and matching efficiency, and show that convenience yields reflect both direct OTC frictions and the interaction between liquidity and return risk. The framework generates closed-form expressions for rates and spreads that facilitate comparative statics and quantitative analysis. Finally, we show that individual investors fail to internalize how their portfolio choices influence aggregate market tightness, which leads to over- or under-investment in liquid assets, depending on the level of tightness and the degree of risk aversion.

Our framework omits several market features that merit further study. First, we do not model large market makers or networks (see e.g., Bech and Atalay, 2010), or the role of collateral, which may shape both the terms and scope of OTC trades. In addition, we abstract from the multi-layered nature of liquidity provision, in which banks provide credit lines and deposits that serve as settlement instruments for non-bank financial institutions. Equipped with suitable data, our framework can advance our understanding of the determinants of convenience yields and, more broadly, asset prices.

References

Acharya, Viral V and Lasse Heje Pedersen, "Asset pricing with liquidity risk," *Journal of financial Economics*, 2005, *77* (2), 375–410.

Aczél, János, *Functional Equations in Several Variables*, Vol. 31 of *Encyclopedia of Mathematics and its Applications*, Cambridge: Cambridge University Press, 1989.

Afonso, Gara and Ricardo Lagos, "The Over-the-Counter Theory of the Fed Funds Market: A Primer," *Journal of Money, Credit and Banking*, June 2015, *47* (2), 127–154.

_ **and** _ , "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 2015, *83* (1), 263–313.

Amador, Manuel and Javier Bianchi, "Bank runs, fragility, and regulation," 2024. NBER Working Paper No. 32341.

Arce, Óscar, Galo Nuño, Dominik Thaler, and Carlos Thomas, "A Large Central Bank Balance Sheet? Floor vs Corridor Systems in a New Keynesian Environment," *Journal of Monetary Economics*, 2019.

Arseneau, David, David Rappoport, and Alexandros Vardoulakis, "Private and public liquidity provision in over-the-counter markets," *Theoretical Economics*, 2020, *15* (4), 1669–1712.

Atkeson, Andrew G., Andrea L. Eisfeldt, and Pierre-Olivier Weill, "Entry and Exit in OTC Derivatives Markets," *Econometrica*, 2015, *83* (6), 2231–2292.

Basak, Suleyman and Domenico Cuoco, "An Equilibrium Model with Restricted Stock Market Participation," *Review of Financial Studies*, 1998, *11* (2), 309–341.

Bech, Morten L. and Enghin Atalay, "The topology of the federal funds market," *Physica A: Statistical Mechanics and its Applications*, 2010, *389* (22), 5223–5246.

Bhattacharya, Sudipto and Douglas Gale, "Preference Shocks, Liquidity, and Central Bank Policy," in William A. Barnett and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*, Cambridge: Cambridge University Press, 1987, pp. 69–88.

Bianchi, Javier, "Overborrowing and systemic externalities in the business cycle," *American Economic Review*, 2011, *101* (7), 3400–3426.

_ **and Saki Bigio**, "Banks, liquidity management, and monetary policy," *Econometrica*, 2022, *90* (1), 391–454.

_ **, _ , and Charles Engel**, "Scrambling for Dollars: International Liquidity, Banks and Exchange Rates," 2023. Mimeo, Minneapolis Fed.

Bigio, Saki and Yuliy Sannikov, "A Model of Intermediation, Money, Interest, and Prices," 2019. Unpublished Manuscript.

Bittner, Christian, Rustam Jamilov, and Farzad Saidi, "Assortative matching, interbank markets, and monetary policy," Technical Report, ECONtribute Discussion Paper 2025.

Burai, Pál, Gergely Kiss, and Patricia Szokol, "Characterization of quasi-arithmetic means without regularity condition," *Acta Mathematica Hungarica*, 2021, *165*, 474–485.

Caballero, Ricardo J and Arvind Krishnamurthy, "International and domestic collateral constraints in a model of emerging market crises," *Journal of monetary Economics*, 2001, *48* (3), 513–548.

Constantinides, George M., "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 1986, *94*, 842–862.

_ **and Darrell Duffie**, "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 1996, *104* (2), 219–240.

Dávila, Eduardo and Anton Korinek, "Pecuniary externalities in economies with financial frictions," *Review of Economic Studies*, 2018, *85* (1), 352–395.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, "Over-the-Counter Markets," *Econometrica*, 2005, *73* (6), 1815–1847.

Farhi, Emmanuel, Mikhail Golosov, and Aleh Tsyvinski, "A Theory of Liquidity and Regulation of Financial Intermediation," *Review of Economic Studies*, 2009, *76* (3), 973–992.

Fiore, Fiorella De, Marie Hoerova, and Harald Uhlig, "Money Markets, Collateral and Monetary Policy," 2018. NBER Working Paper 25319.

Geanakoplos, John and Kieran James Walsh, "Inefficient Liquidity Provision," *Economic Theory*, 2017. Advance online publication.

He, Zhiguo and Konstantin Milbradt, "Endogenous liquidity and defaultable bonds," *Econometrica*, 2014, *82* (4), 1443–1508.

Holmstrom, Bengt and Jean Tirole, "LAPM: A Liquidity Based Asset Pricing Model," *Journal of Finance*, October 2001, *56* (5), 1837–1867.

Hosios, Arthur J, "On the efficiency of matching and related models of search and unemployment," *The Review of Economic Studies*, 1990, *57* (2), 279–298.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, "Heterogeneity in decentralized asset markets," *Theoretical Economics*, 2022, *17* (3), 1313–1356.

_ , _ , and _ , *The Economics of Over-the-Counter Markets: A Toolkit for the Analysis of Decentralized Exchange*, Princeton University Press, 2025.

Ismail, Ali Haider and Diego Zuniga, "Welfare Cost of Convenience Yield," 2025. Working Paper.

Jacklin, Charles J., "Demand Deposits, Trading Restrictions, and Risk Sharing," in Edward C. Prescott and Neil Wallace, eds., *Contractual Arrangements for Intertemporal Trade*, Minneapolis: University of Minnesota Press, 1987, pp. 26–47.

Kiyotaki, Nobuhiro and Randall Wright, "A search-theoretic approach to monetary economics," *The American Economic Review*, 1993, pp. 63–77.

Krishnamurthy, Arvind and Annette Vissing-Jørgensen, "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 2012, *120* (2), 233–267.

Krueger, Dirk and Hanno Lustig, "When is market incompleteness irrelevant for the price of aggregate risk (and when is it not)?," *Journal of Economic Theory*, 2010, *145* (1), 1–41.

Lagos, Ricardo, "Asset prices and liquidity in an exchange economy," *Journal of Monetary Economics*, 2010, *57* (8), 913–930.

_ and Guillaume Rocheteau, "Liquidity in Asset Markets with Search Frictions," *Econometrica*, 2009, *77*, 403–426.

_ **and Randall Wright**, "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 2005.

_ **, Guillaume Rocheteau, and Randall Wright**, "Liquidity: A New Monetarist Perspective," *Journal of Economic Literature*, 2017, *55* (2), 371–440.

Lenel, Moritz, Monika Piazzesi, and Martin Schneider, "The short rate disconnect in a monetary economy," *Journal of Monetary Economics*, 2019, *106*, 59–77.

Lester, Benjamin, Andrew Postlewaite, and Randall Wright, "Information, Liquidity, Asset Prices, and Monetary Policy," *Review of Economic Studies*, 2012, *79* (3), 1209–1238.

Lopez-Salido, David and Annette Vissing-Jorgenson, "Reserve Demand, Interest Rate Control, and Quantitative Tightening," 2023. Working Paper.

Lorenzoni, Guido, "Inefficient Credit Booms," *Review of Economic Studies*, July 2008, *75* (2), 809–833.

Nagel, Stefan, "The Liquidity Premium of Near-Money Assets," *The Quarterly Journal of Economics*, 2016, *131* (4), 1927–1971.

Piazzesi, Monika, Ciaran Rogers, and Martin Schneider, "Money and Banking in a New Keynesian Model," 2019. Working Paper, Stanford.

Poole, William, "Commercial bank reserve management in a stochastic model: implications for monetary policy," *The Journal of Finance*, 1968, *23* (5), 769–791.

Ríos-Rull, José-Víctor, Joao Ritto, Tamon Takamura, and Yaz Terajima, "Real Monetary Policy," October 2024. Mimeo, University of Pennsylvania.

Shi, Shouyong, "A Divisible Search Model of Fiat Money," *Econometrica*, 1997, *65*, 75–102.

Silva, Dejanir, Juan Passadore, and Mahyar Kargar, "Liquidity and Risk in OTC Markets: A Theory of Asset Pricing and Portfolio Flows," 2023. Working Paper.

Uslu, Semih, "Pricing and Liquidity in Decentralized Asset Markets," *Econometrica*, 2019, *87* (6), 2079–2140.

Vayanos, Dimitri and Jean-Luc Vila, "Equilibrium Interest Rate and Liquidity Premium with Transaction Costs," *Economic Theory*, 1999, *13*, 509–539.

Williamson, Stephen D. and Randall Wright, "New Monetarist Economics: Models," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Elsevier, 2010, chapter 2, p. 25–96.

Wong, Russell and Mengbo Zhang, "Disintermediating the Federal Funds Market," 2023. Working Paper.

Yared, Pierre, "Public Debt Under Limited Private Credit," *Journal of the European Economic Association*, 2013, *11* (2), 229–245.

A. Proofs

A.1 Proof of Proposition 1

We begin with an auxiliary lemma showing that market tightness follows a difference equation. With the market tightness, we obtain the matching probabilities at each round.

Lemma 1. *Let θ_0 be the initial market tightness. Then, the ratio $\{\theta_n\}$ features the following law of motion:*

$$\theta_n = \theta_{n-1}\frac{(1-\lambda_N G(1/\theta_{n-1},1))}{(1-\lambda_N G(1,\theta_{n-1}))} \qquad \forall n \in \{1,2,\ldots,N\}$$

and the matching probabilities can be expressed in terms of the ratio via:

$$\psi_n^+ = \lambda_N G(1,\theta_{n-1}) \text{ and } \psi_n^- = \lambda_N G(1/\theta_{n-1},1).$$

Proof. By definition and homogeneity:

$$\theta_n = \frac{S_n^-}{S_n^+} = \frac{S_{n-1}^- - z_n}{S_{n-1}^+ - z_n} = \theta_{n-1}\frac{(1-\lambda_N G(1/\theta_{n-1},1))}{(1-\lambda_N G(1,\theta_{n-1}))}, \qquad \forall n \in \{1,2,\ldots,N\}.$$

where the second equality follows from the definition of z_n and uses its homogeneity property. □

The lemma shows that we can track matching probabilities in terms of the initial market tightness, without reference to the terms of trade. It also shows that these probabilities are scale invariant. We use these observations in what follows, as the Lemma permits us to treat trading probabilities as exogenous series.

The proof also makes use of the following standard result in probability, which we include for ease of completeness:

Lemma 2 (Conditional Probability Decomposition)**.** *For any starting round $n \geq 0$, the conditional probabilities of matching satisfy:*

$$\sum_{k=n+1}^{N} \psi_k^{\pm} \prod_{m=n+1}^{k-1}(1-\psi_m^{\pm}) + \prod_{m=n+1}^{N}(1-\psi_m^{\pm}) = 1.$$

That is, starting from round n, the probability of matching in some future round plus the probability of never matching equals 1.

Proof. Let $P_{n,j} = \prod_{m=n+1}^{j}(1-\psi_m^{\pm})$ be the probability of not matching from round $n+1$ through round j, with the convention that $P_{n,n} = 1$.

Then,

$$\begin{aligned}\sum_{k=n+1}^{N} \psi_k^{\pm} \prod_{m=n+1}^{k-1} (1-\psi_m^{\pm}) &= \sum_{k=n+1}^{N} \psi_k^{\pm} P_{n,k-1} \\ &= \sum_{k=n+1}^{N} [P_{n,k-1} - P_{n,k}] \quad \text{(since } P_{n,k} = P_{n,k-1}(1-\psi_k^{\pm})\text{)} \\ &= (P_{n,n} - P_{n,n+1}) + (P_{n,n+1} - P_{n,n+2}) + \cdots + (P_{n,N-1} - P_{n,N}) \\ &= P_{n,n} - P_{n,N} \\ &= 1 - \prod_{m=n+1}^{N} (1-\psi_m^{\pm})\end{aligned}$$

where we used $P_{n,n} = 1$ and the telescoping sum. The result follows by rearranging. □

Next, we describe the limit of the bargaining problem as $\Delta \to 0$. Recall that the trader's estimate of equity, excluding its own trade, is:

$$\mathcal{E}^j(\Delta) \equiv \sum_{i \in \mathbb{I}} a_{t+1}^i R_{t+1}^i + m_{t+1} R_{t+1}^m + \chi_{t+1}(s^j - \text{sign}(s^j)\Delta).$$

The proof of Proposition 1 is as follows.

Proof. The Nash bargaining problem at round n with trade size Δ has surplus for trader with position s^j:

$$\mathcal{S}_n^{\text{sign}(s^j)}(\Delta) = V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)(r_n^f - r^m)\Delta) - J_U^{\text{sign}(s^j)}(n;\Delta).$$

Step 1: Outside option recursion. The unmatched value satisfies:

$$J_U^{\text{sign}(s^j)}(n;\Delta) = \psi_{n+1}^{\text{sign}(s^j)} V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)(r_{n+1}^f - r^m)\Delta) + (1 - \psi_{n+1}^{\text{sign}(s^j)}) J_U^{\text{sign}(s^j)}(n+1;\Delta),$$

with terminal round, N, values given by

$$\begin{aligned}J_U^+(N;\Delta) &= V(\mathcal{E}^j(\Delta)) \quad \text{(surplus traders hold cash at rate } r^m\text{)}, \\ J_U^-(N;\Delta) &= V(\mathcal{E}^j(\Delta) - (r^w - r^m)\Delta) \quad \text{(deficit traders borrow at rate } r^w\text{)}.\end{aligned}$$

Expanding the recursion forward from round n:

$$J_U^{\text{sign}(s^j)}(n;\Delta) = \sum_{k=n+1}^{N} \left[\prod_{m=n+1}^{k-1} (1-\psi_m^{\text{sign}(s^j)})\right] \psi_k^{\text{sign}(s^j)} V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)(r_k^f - r^m)\Delta) + \left[\prod_{m=n+1}^{N} (1-\psi_m^{\text{sign}(s^j)})\right] V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)\chi_N^{\text{sign}(s^j)}\Delta).$$

By Lemma 2, this corresponds to the value of $\mathbb{E}\left[V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)(r_n^f - r^m)\Delta) | \text{unmatched by n}\right]$. Importantly, notice that the expectation assumes that the negotiated rate r_n^f depends only on future rounds, an assumption that we verify below.

Now, consider the normalized difference:

$$\frac{J_U^{\text{sign}(s^j)}(n;\Delta) - V(\mathcal{E}^j(\Delta))}{\Delta} = \sum_{k=n+1}^{N} \left[\prod_{m=n+1}^{k-1} (1-\psi_m^{\text{sign}(s^j)})\right] \psi_k^{\text{sign}(s^j)} \frac{V(\mathcal{E}^j + \text{sign}(s^j)(r_k^f - r^m)\Delta) - V(\mathcal{E}^j)}{\Delta} + \left[\prod_{m=n+1}^{N} (1-\psi_m^{\text{sign}(s^j)})\right] \frac{V(\mathcal{E}^j + \text{sign}(s^j)\chi_{N+1}^{\text{sign}(s^j)}\Delta) - V(\mathcal{E}^j)}{\Delta}.$$

As $\Delta \to 0$, by the definition of the derivative,

$$\lim_{\Delta\to 0} \frac{J_U^{\text{sign}(s^j)}(n;\Delta) - V(\mathcal{E}^j(\Delta))}{\Delta} = \text{sign}(s^j) \cdot V'(\mathcal{E}^j) \cdot \chi_n^{\text{sign}(s^j)},$$

where

$$\chi_n^{\text{sign}(s^j)} = \sum_{k=n+1}^{N} (r_k^f - r^m) \left[\prod_{m=n+1}^{k-1} (1-\psi_m^{\text{sign}(s^j)})\right] \psi_k^{\text{sign}(s^j)} + \chi_{N+1}^{\text{sign}(s^j)} \prod_{m=n+1}^{N} (1-\psi_m^{\text{sign}(s^j)}). \quad \text{(A.1)}$$

The variable $\chi_n^{\text{sign}(s^j)}$ represents the expected financing cost/benefit conditional on being unmatched at round n, for $+$ and $-$ positions. The variables satisfy the following recursion:

$$\chi_n^{\text{sign}(s^j)} = \psi_{n+1}^{\text{sign}(s^j)}(r_{n+1}^f - r^m) + (1-\psi_{n+1}^{\text{sign}(s^j)})\chi_{n+1}^{\text{sign}(s^j)},$$

with terminal conditions $\chi_N^+ = 0$ and $\chi_N^- = r^w - r^m$.

Step 2: Limiting surplus. Consider the following term:

$$\frac{S_n^{\text{sign}(s^j)}(\Delta)}{\Delta} = \frac{\left(V(\mathcal{E}^j(\Delta) + \text{sign}(s^j)(r_n^f - r^m)\Delta) - V(\mathcal{E}^j(\Delta))\right) - \left(J_U^{\text{sign}(s^j)}(n;\Delta) - V(\mathcal{E}^j(\Delta))\right)}{\Delta}.$$

Taking the $\Delta \to 0$ limit and substituting the result in Step 1

$$\lim_{\Delta \to 0} \frac{\mathcal{S}_n^{\text{sign}(s^j)}(\Delta)}{\Delta} = \text{sign}(s^j) V'(\mathcal{E}^j(0))[(r_n^f - r^m) - \chi_n^{\text{sign}(s^j)}].$$

Step 3: Nash bargaining. The bargained rate solves:

$$r_n^f(\Delta) = \arg\max_{r_n} [S_n^-(\Delta)]^\eta [S_n^+(\Delta)]^{1-\eta}.$$

Since multiplying the objective by a positive constant Δ^{-1} doesn't change the maximizer

$$r_n^f(\Delta) = \arg\max_{r_n} \frac{[S_n^-(\Delta)]^\eta [S_n^+(\Delta)]^{1-\eta}}{\Delta} = \arg\max_{r_n} \left[\frac{S_n^-(\Delta)}{\Delta}\right]^\eta \left[\frac{S_n^+(\Delta)}{\Delta}\right]^{1-\eta}.$$

Taking the limit as $\Delta \to 0$:

$$r_n^f = \lim_{\Delta \to 0} r_n^f(\Delta) = \lim_{\Delta \to 0} \left\{ \arg\max_{r_n} \left[\frac{S_n^-(\Delta)}{\Delta}\right]^\eta \left[\frac{S_n^+(\Delta)}{\Delta}\right]^{1-\eta} \right\}.$$

By the theorem of the maximum, since the objective function is continuous in both r_n and Δ and the constraint set $[r^m, r^w]$ is bounded, we can pass limits inside the maximum operator:

$$\lim_{\Delta \to 0} r_n^f = \arg\max_{r_n} \left\{ \lim_{\Delta \to 0} \left[\frac{S_n^-(\Delta)}{\Delta}\right]^\eta \left[\frac{S_n^+(\Delta)}{\Delta}\right]^{1-\eta} \right\}.$$

Substituting the limits from Step 2, we obtain

$$r_n^f = \arg\max_{r_n} \left\{ [V'(\mathcal{E}^j(0))]^\eta [V'(\mathcal{E}^k(0))]^{1-\eta} \cdot [\chi_n^- - (r_n - r^m)]^\eta [(r_n - r^m) - \chi_n^+]^{1-\eta} \right\}.$$

Since $V'(\mathcal{E}^j(0))$ and $V'(\mathcal{E}^k(0))$ are positive constants independent of r_n^f, this reduces to:

$$r_n^f = \arg\max_{r_n} [\chi_n^- - (r_n - r^m)]^\eta [(r_n - r^m) - \chi_n^+]^{1-\eta},$$

where χ_n^+ and χ_n^- satisfy A.1.

Step 4: First-order condition. We now solve the maximization problem. The FOC yields: $\eta[(r_n^f - r^m) - \chi_n^+] = (1-\eta)[\chi_n^- - (r_n^f - r^m)]$. Therefore,

$$r_n^f = r^m + (1-\eta)\chi_n^- + \eta\chi_n^+,$$

as stated by the proposition. This justifies the assumption that outside options depend only on the number of rounds. □

A.2 Proof of Corollary 1

Proof. The matching probabilities over N rounds are

$$\Psi^+ = 1 - \prod_{n=1}^{N}(1-\psi_n^+) \quad \text{and} \quad \Psi^- = 1 - \prod_{n=1}^{N}(1-\psi_n^-).$$

These represent the probability of matching at least once during the OTC stage.

Verification of χ^+ and χ^-. We have the recursion (A.1). Solving that recursion forward, from round 0 to round N, yields

$$\chi_0^\pm = \sum_{n=1}^{N}(r_n^f - r^m)\psi_n^\pm \prod_{k=1}^{n-1}(1-\psi_k^\pm) + \chi_{N+1}^\pm \prod_{n=1}^{N}(1-\psi_n^\pm). \tag{A.2}$$

where we used $\chi_{N+1}^+ = 0$ and $\chi_{N+1}^- = r^w - r^m$. The first term is the OTC rates weighted by the unconditional matching probabilities, and the second term is the terminal value weighted by the probability of never matching, $(1-\Psi^\pm)$.

Recall that probabilities in the expansion add up to 1: apply Lemma 2 to $n = 1$. Thus, we can define the volume-weighted average rate as:

$$\overline{r}^f - r^m = \sum_{n=1}^{N}(r_n^f - r^m)\frac{\psi_n^\pm \prod_{k=1}^{n-1}(1-\psi_k^\pm)}{\Psi^\pm},$$

where the weights in the sum add up to 1. Substituting the definition of weighted rates in (A.2), we obtain:

$$\begin{aligned}
\chi_0^+ &= \Psi^+(\overline{r}^f - r^m) + 0 \cdot (1-\Psi^+) = \Psi^+(\overline{r}^f - r^m) = \chi^+, \\
\chi_0^- &= \Psi^-(\overline{r}^f - r^m) + (r^w - r^m)(1-\Psi^-) = \chi^-.
\end{aligned}$$

Consistency of Weights. Probabilities are proportional to traded amounts since $z_n = \psi_n^\pm S_{n-1}^\pm$ and $S_{n-1}^\pm = \prod_{k=1}^{n-1}(1-\psi_k^\pm)S_0^\pm$ and $\sum_{n=1}^{N} z_n = \Psi^\pm S_0^\pm$. Thus, the weights $\varkappa_n^\pm = \frac{\psi_n^\pm \prod_{k=1}^{n-1}(1-\psi_k^\pm)}{\Psi^\pm}$ are identical for surplus and deficit sides because of the homogeneity of the matching function (see Lemma 1), ensuring a unique $\overline{r}^f$. □

A.3 Proof of Proposition 2

Consider the continuous-time limit as $N \to \infty$. Let $\Delta = 1/N$ denote the time between rounds, and index continuous time by $\tau \in [0, 1]$. We derive the ODE in the statement of the Lemma.

Proof. **Step 1: Evolution of masses.** In discrete time, the masses of surplus and deficit positions evolve as:

$$S^{\pm}_{n+1} = S^{\pm}_n - \lambda_N G(S^+_n, S^-_n).$$

With $\lambda_N = \lambda/N$ and continuous time $\tau = n/N$, taking $\Delta \to 0$,

$$\dot{S}^{\pm}_\tau = \lim_{\Delta \to 0} \frac{S^{\pm}_{\tau+\Delta} - S^{\pm}_\tau}{\Delta} = -\lambda G(S^+_\tau, S^-_\tau).$$

Both sides shrink at the same absolute rate since matches clear equal amounts from each side.

Step 2: Dynamics of tightness. Define $\theta_\tau = S^-_\tau / S^+_\tau$. Taking logarithms and differentiating:

$$\begin{aligned}
\frac{\dot{\theta}_\tau}{\theta_\tau} &= \frac{d}{d\tau} \ln(\theta_\tau) = \frac{d}{d\tau}[\ln(S^-_\tau) - \ln(S^+_\tau)] \\
&= \frac{\dot{S}^-_\tau}{S^-_\tau} - \frac{\dot{S}^+_\tau}{S^+_\tau} \\
&= -\lambda \frac{G(S^+_\tau, S^-_\tau)}{S^-_\tau} + \lambda \frac{G(S^+_\tau, S^-_\tau)}{S^+_\tau}.
\end{aligned}$$

Using homogeneity of degree one, we obtain $G(S^+, S^-) = S^+ G(1, \theta) = S^- G(\theta^{-1}, 1)$. Therefore,

$$\frac{\dot{\theta}_\tau}{\theta_\tau} = -\lambda(\theta_\tau^{-1}, 1) + \lambda(1, \theta_\tau).$$

Using the normalized matching function, $\gamma(\theta) \equiv G(1, \theta)$, and noting that $G(\theta^{-1}, 1) = \gamma(\theta^{-1})$ by definition,

$$\dot{\theta}_\tau = -\lambda \theta_\tau [\gamma(\theta_\tau^{-1}) - \gamma(\theta_\tau)].$$

Step 3: Matching intensities. The instantaneous matching rates per unit mass are the limit as $\Delta \to 0$ of $\psi^{\pm}_n / \Delta = \lambda_n \frac{G(S^+_n, S^-_n)}{S^{\pm}_n \Delta}$.[38] By the calculations above,

$$\psi^+_\tau = \lambda \gamma(\theta_\tau) \quad \text{and} \quad \psi^-_\tau = \lambda \gamma(\theta_\tau^{-1}).$$

Note that $\psi^+_\tau = \theta_\tau \psi^-_\tau$ follows from $\gamma(\theta) = \theta \gamma(\theta^{-1})$ (by symmetry). □

[38] After dividing by Δ we interpret $\psi^{\pm}_\tau$ as an intensity.

A.4 Proof of Corollary 2

Proof. From Lemma 4.2, the ODE for market tightness is,

$$\dot{\theta}_\tau = -\lambda\theta_\tau[\gamma(\theta_\tau^{-1}) - \gamma(\theta_\tau)]. \tag{A.3}$$

1. Steady state. If $\theta_0 = 1$, then $\gamma(\theta_0^{-1}) = \gamma(1) = \gamma(\theta_0)$, so $\dot{\theta}_0 = 0$. Then, by uniqueness of solutions to ODEs, $\theta_\tau = 1$ for all $\tau \in [0, 1]$.

2. Monotonicity. The sign of $\dot{\theta}_\tau$ depends on $\gamma(\theta_\tau^{-1}) - \gamma(\theta_\tau)$. By symmetry of G,

$$\gamma(\theta^{-1}) = G(1, \theta^{-1}) = G(\theta^{-1}, 1) = \frac{G(1,\theta)}{\theta} = \frac{\gamma(\theta)}{\theta}.$$

Therefore,

$$\dot{\theta}_\tau = -\lambda\theta_\tau \left[\frac{\gamma(\theta_\tau)}{\theta_\tau} - \gamma(\theta_\tau)\right] = \lambda\gamma(\theta_\tau)(\theta_\tau - 1)$$

Since $\gamma(\theta) > 0$ for all $\theta > 0$: (i) If $\theta_0 > 1$, then $\dot{\theta}_0 > 0$, so θ_τ increases. (ii) If $\theta_0 < 1$, then $\dot{\theta}_0 < 0$, so θ_τ decreases.

3. Effect of matching functions. Consider two matching functions $\gamma, \tilde{\gamma}$, such that $\gamma(\theta) < \tilde{\gamma}(\theta)$ for all . From the ODE (A.3): (i) If $\theta > 1$: $\dot{\theta}$ is larger under $\tilde{\gamma}$, so θ rises faster. (ii) If $\theta < 1$: $\dot{\theta}$ is smaller under $\tilde{\gamma}$, so θ falls faster. □

A.5 Proof of Corollary 3

We proceed to derive the integral form in the proposition.

Proof. **Step 1: Continuous-time limit of the recursions.** From the discrete-rounds recursion:

$$\chi_n^\pm = \psi_{n+1}^\pm(r_{n+1}^f - r^m) + (1 - \psi_{n+1}^\pm)\chi_{n+1}^\pm$$

we can rearrange terms to obtain $\chi_{n+1}^\pm - \chi_n^\pm = \psi_{n+1}^\pm[\chi_{n+1}^\pm - (r_{n+1}^f - r^m)]$. Dividing both sides by Δ, substituting the discrete round probabilities for the corresponding intensities, and the change of variables $\Delta = 1/N$ and $\tau = n/N$, we have

$$\frac{\chi_{\tau+\Delta}^\pm - \chi_\tau^\pm}{\Delta} = \frac{\psi_{\tau+\Delta}^\pm}{\Delta}[\chi_{\tau+\Delta}^\pm - (r_{\tau+\Delta}^f - r^m)].$$

Now, recall that,

$$\lim_{\Delta\to 0}\frac{\psi^{\pm}_{\tau+\Delta}}{\Delta}=\lim_{\Delta\to 0}\frac{\lambda_N(\Delta)}{\Delta}\cdot\frac{G(S^+_n,S^-_n)}{S^{\pm}_\tau}=\lambda\frac{G(S^+_\tau,S^-_\tau)}{S^{\pm}_\tau}.$$

Thus, taking $\Delta\to 0$, we obtain an ODE for the outside options,

$$\dot{\chi}^{\pm}_\tau=\psi^{\pm}_\tau[\chi^{\pm}_\tau-(r^f_\tau-r^m)].$$

Step 2: Substituting the bargained rate. After the change of index from rounds to time

$$r^f_\tau=r^m+(1-\eta)\chi^-_\tau+\eta\chi^+_\tau.$$

Hence, the ODE for the outside options becomes a coupled system,

$$\dot{\chi}^+_\tau=\psi^+_\tau[\chi^+_\tau-(1-\eta)\chi^-_\tau-\eta\chi^+_\tau]=-(1-\eta)\psi^+_\tau(\chi^-_\tau-\chi^+_\tau) \tag{A.4}$$

$$\dot{\chi}^-_\tau=\psi^-_\tau[\chi^-_\tau-(1-\eta)\chi^-_\tau-\eta\chi^+_\tau]=\eta\psi^-_\tau(\chi^-_\tau-\chi^+_\tau). \tag{A.5}$$

Step 3: Solving for the joint surplus. Define $\Sigma_\tau\equiv\chi^-_\tau-\chi^+_\tau$ as the joint surplus. Then, the joint surplus must satisfy the following ODE,

$$\dot{\Sigma}_\tau=\dot{\chi}^-_\tau-\dot{\chi}^+_\tau=[\eta\psi^-_\tau+(1-\eta)\psi^+_\tau]\Sigma_\tau.$$

This is a linear ODE with backward solution,

$$\Sigma_\tau=\Sigma_1\exp\left(-\int_\tau^1[\eta\psi^-_s+(1-\eta)\psi^+_s]ds\right).$$

Since the terminal conditions for the outside options are $\chi^+_1=0$ and $\chi^-_1=r^w-r^m$, the terminal condition for the ODE is given by $\Sigma_1=r^w-r^m$. Thus,

$$\Sigma_\tau=(r^w-r^m)\exp\left(-\int_\tau^1[\eta\psi^-_s+(1-\eta)\psi^+_s]ds\right). \tag{A.6}$$

Step 4: Solving for individual components. Substituting the integral in (A.6) back into the ODE system (A.4)-(A.5):

$$\dot{\chi}^+_\tau=-(1-\eta)\psi^+_\tau\Sigma_\tau,\qquad\dot{\chi}^-_\tau=\eta\psi^-_\tau\Sigma_\tau.$$

Integrating from τ to 1 and using the terminal conditions,

$$\chi_\tau^+ - 0 = \int_\tau^1 (1-\eta)\psi_y^+ \Sigma_y dy, \quad \chi_\tau^- - (r^w - r^m) = -\int_\tau^1 \eta\psi_y^- \Sigma_y dy.$$

Substituting (A.6), we obtain the expression for Σ_y,

$$\chi_\tau^+ = (r^w - r^m)\int_\tau^1 (1-\eta)\psi_y^+ \exp\left(-\int_y^1 [\eta\psi_x^- + (1-\eta)\psi_x^+]dx\right) dy,$$
$$\chi_\tau^- = (r^w - r^m)\left[1 - \int_\tau^1 \eta\psi_y^- \exp\left(-\int_y^1 [\eta\psi_x^- + (1-\eta)\psi_x^+]dx\right) dy\right].$$

The convenience-yield coefficients are defined as the $\tau = 0$ values of the integrals above: $\chi^\pm = \chi_0^\pm$, and the OTC rate is $r_\tau^f = r^m + (1-\eta)\chi_\tau^- + \eta\chi_\tau^+$. □

To verify that our solution is consistent with the discrete case in that weighted-average rates can be obtain from $\chi_0^\pm$, we trace the relationship between matching intensities and probabilities in continuous time. The matching intensities $\psi_\tau^\pm$ represent the instantaneous matching rate per unit time. These are intensities, but cannot be interpreted directly as PDFs. The probability of not matching from time 0 to τ is therefore,

$$1 - F^\pm(\tau) = \exp\left(-\int_0^\tau \psi_s^\pm ds\right),$$

where the CDF of matching by time τ is $F^\pm(\tau) \equiv 1 - \exp\left(-\int_0^\tau \psi_s^\pm ds\right)$. Hence, the PDF (probability density of matching at time τ) is,

$$f^\pm(\tau) = \frac{dF^\pm}{d\tau} = \psi_\tau^\pm \exp\left(-\int_0^\tau \psi_s^\pm ds\right) = \psi_\tau^\pm (1 - F^\pm(\tau)).$$

This confirms the hazard rate relationship, $\psi_\tau^\pm = f^\pm(\tau)/(1-F^\pm(\tau))$. The volume-weighted average rate is $\overline{r}^f = \int_0^1 \varkappa_\tau^\pm r_\tau^f d\tau$, where the weights $\varkappa_\tau^\pm = f^\pm(\tau)/\Psi^\pm$ represent the fraction of total volume traded at time τ.

Verification: Integrating the ODE $\dot{\chi}_\tau^- = \psi_\tau^-[\chi_\tau^- - (r_\tau^f - r^m)]$ multiplied by $(1 - F^-(\tau))$ and using integration by parts, we obtain

$$\Psi^-(\overline{r}^f - r^m) = \chi_0^- - (1-\Psi^-)(r^w - r^m)$$

which confirms $\chi_0^- = \Psi^-(\overline{r}^f - r^m) + (1-\Psi^-)(r^w - r^m) = \chi^-$. Similarly for $\chi_0^+ = \chi^+$. This is the same consistency condition as in the discrete rounds.

A.6 Proof of Proposition 3

Proof. We begin establishing the result for χ^+. From Proposition 3, we obtained (14):

$$\chi_\tau^+ = \int_\tau^1 (1-\eta)\psi_t^+ \Sigma_t dt,$$

where the surplus satisfies:

$$\Sigma_\tau = (r^w - r^m) \exp\left(- \int_\tau^1 [\eta \psi_x^- + (1-\eta)\psi_x^+] dx \right).$$

Recall that overall matching probabilities, given in (11), satisfy:

$$e^{-\int_\tau^1 \psi_s^+ ds} = 1 - \Psi_\tau^+, \quad e^{-\int_\tau^1 \psi_s^- ds} = 1 - \Psi_\tau^-.$$

Thus, we have that

$$\Sigma_\tau = (r^w - r^m)(1-\Psi_\tau^+)^{1-\eta}(1-\Psi_\tau^-)^\eta. \tag{A.7}$$

Consider the following proposed solution with χ_τ^+ as a function of θ_τ:

$$\chi^+(\theta_\tau) = (r^w - r^m)\frac{\bar\theta - \bar\theta^\eta \theta_\tau^{1-\eta}}{\bar\theta - 1}.$$

If the proposal is correct, its time derivative must coincide with the term inside the integral, and we must verify that it satisfies the terminal condition.[39] Taking the derivative with respect to τ:

$$\frac{\partial \chi_\tau^+}{\partial \tau} = (r^w - r^m)\frac{-(1-\eta)\bar\theta^\eta \theta^{-\eta}}{\bar\theta - 1}\dot\theta. \tag{A.8}$$

By market clearing

$$\bar\theta = \frac{1-\Psi_\tau^-}{1-\Psi_\tau^+}\theta_\tau, \tag{A.9}$$

a condition we can use to substitute out $\bar\theta$ from (A.8). We obtain that

$$\frac{\partial \chi_\tau^+}{\partial \tau} = -(r^w - r^m)(1-\eta)\frac{\left(\frac{1-\Psi_\tau^-}{1-\Psi_\tau^+}\right)^\eta}{\frac{1-\Psi_\tau^-}{1-\Psi_\tau^+}\theta - 1}\dot\theta.$$

[39] An alternative proof that does not involve guess and verify can be obtained by separation of variables. We thank Chenming Zhang for this observation.

Using the market clearing condition $\Psi_\tau^- \theta_\tau = \Psi_\tau^+$, the denominator in the fraction becomes

$$\frac{\theta - \Psi_\tau^- \theta - 1 + \Psi_\tau^+}{1 - \Psi_\tau^+} = \frac{\theta - 1}{1 - \Psi_\tau^+}.$$

We then have that

$$\frac{\partial \chi_\tau^+}{\partial \tau} = -(r^w - r^m)(1-\eta)\frac{(1-\Psi_\tau^-)^\eta (1-\Psi_\tau^+)^{1-\eta}}{\theta - 1}\dot{\theta}. \tag{A.10}$$

Next, we substitute out $\dot{\theta}$. Recall from equation (13) that market tightness follows

$$\dot{\theta} = -\lambda\theta(\gamma(\theta^{-1}) - \gamma(\theta)).$$

Replacing the matching intensities given in (10), $\psi_\tau^+ = \lambda\gamma(\theta_\tau)$ and $\psi_\tau^- = \lambda\gamma(\theta_\tau^{-1})$, combined with the market clearing condition $\psi_\tau^- \theta_\tau = \psi_\tau^+$ we arrive at:

$$\dot{\theta} = -(\lambda\theta\gamma(\theta^{-1}) - \lambda\theta\gamma(\theta)) = \psi_\tau^+(\theta - 1).$$

Substituting this step into (A.10), we obtain

$$\frac{\partial \chi_\tau^+}{\partial \tau} = -(r^w - r^m)(1-\eta)(1-\Psi_\tau^-)^\eta (1-\Psi_\tau^+)^{1-\eta}\psi_\tau^+ = -(1-\eta)\psi_\tau^+ \Sigma_\tau.$$

This matches the integrand exactly. Since the terminal condition is also satisfied (at $\tau = 1$, $\theta_1 = \bar{\theta}$ giving $\chi_1^+ = 0$), the formula is verified. Applying the integral to this term yields χ^+. The result is valid for any matching function.

To verify the formula,

$$\chi^- = (r^w - r^m)\frac{\bar{\theta} - \bar{\theta}^\eta \theta_0^{-\eta}}{\bar{\theta} - 1},$$

we use the property that $\Sigma_\tau = \chi_\tau^- - \chi_\tau^+$. Thus, we obtain that if the formula is correct:

$$\Sigma_\tau = (r^w - r^m)\frac{\bar{\theta}^\eta \theta_\tau^{1-\eta} - \bar{\theta}^\eta \theta_\tau^{-\eta}}{\bar{\theta} - 1} = (r^w - r^m)\bar{\theta}^\eta \theta_\tau^{1-\eta}\frac{\theta_\tau - 1}{\theta_\tau(\bar{\theta} - 1)}. \tag{A.11}$$

We now verify this formula using (A.7):

$$\Sigma_\tau = (r^w - r^m)(1 - \Psi_\tau^+)\left(\frac{1-\Psi_\tau^-}{1-\Psi_\tau^+}\right)^\eta.$$

Equation (A.9) implies

$$\bar{\theta} - 1 = \frac{1 - \Psi_\tau^-}{1 - \Psi_\tau^+}\theta - 1 = \frac{\theta - 1}{1 - \Psi_\tau^+} \Rightarrow 1 - \Psi_\tau^+ = \frac{\theta - 1}{\bar{\theta} - 1}. \tag{A.12}$$

where we used the clearing condition. Back into the equation above, we obtain

$$\Sigma_\tau = (r^w - r^m)\left(\frac{\theta - 1}{\bar{\theta} - 1}\right)\left(\frac{\bar{\theta}}{\theta}\right)^\eta.$$

Multiplying and dividing by θ yields the target (A.11).

□

A.7 Proof of Corollary 4

Proof. **Part 1: Monotonicity of χ^+ and χ^-.** Recall that the outside option for surplus positions is

$$\chi_\tau^+ = (r^w - r^m)\left[\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{1-\eta}}{\bar{\theta} - 1}\right].$$

Differentiating with respect to τ and noting that $\bar{\theta}$ is fixed we obtain

$$\frac{d\chi_\tau^+}{d\tau} = -\frac{(r^w - r^m)(1-\eta)\bar{\theta}^\eta}{(\bar{\theta} - 1)} \cdot \frac{\dot{\theta}_\tau}{\theta_\tau^\eta}.$$

For $\theta_0 > 1$ all terms are positive, yielding $\frac{d\chi_\tau^+}{d\tau} < 0$. For $\theta_0 < 1$ the denominator has $\bar{\theta} - 1 < 0$ and $\dot{\theta}_\tau < 0$, so their ratio is positive, again yielding $\frac{d\chi_\tau^+}{d\tau} < 0$.

Likewise

$$\chi_\tau^- = (r^w - r^m)\left[\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{-\eta}}{\bar{\theta} - 1}\right].$$

Then, differentiating,

$$\frac{d\chi_\tau^-}{d\tau} = \frac{(r^w - r^m)\eta\bar{\theta}^\eta}{(\bar{\theta} - 1)} \cdot \frac{\dot{\theta}_\tau}{\theta_\tau^{\eta+1}}.$$

For both $\theta_0 > 1$ and $\theta_0 < 1$, the ratio $\frac{\dot{\theta}_\tau}{(\bar{\theta}-1)}$ is positive, yielding $\frac{d\chi_\tau^-}{d\tau} > 0$.

Part 3: Monotonicity of r_τ^f. From Proposition 1, the OTC rate satisfies $r_\tau^f = r^m + \eta\chi_\tau^+ + (1-\eta)\chi_\tau^-$. Differentiating and substituting Parts 1 and 2

$$\frac{dr_\tau^f}{d\tau} = \frac{(r^w - r^m)\eta(1-\eta)\bar{\theta}^\eta \dot{\theta}_\tau}{(\bar{\theta} - 1)\theta_\tau^{\eta+1}}(1 - \theta_\tau).$$

The sign is determined by $(1-\theta_\tau)\cdot\frac{\dot{\theta}_\tau}{(\bar{\theta}-1)}$:

- If $\theta_0 > 1$: then $(1-\theta_\tau) < 0$ and $\frac{\dot{\theta}_\tau}{(\bar{\theta}-1)} > 0$, so $\frac{dr^f_\tau}{d\tau} < 0$.
- If $\theta_0 < 1$: then $(1-\theta_\tau) > 0$ and $\frac{\dot{\theta}_\tau}{(\bar{\theta}-1)} > 0$, so $\frac{dr^f_\tau}{d\tau} > 0$.
- If $\theta_0 = 1$: then $(1-\theta_\tau) = 0$, so $\frac{dr^f_\tau}{d\tau} = 0$.

This completes the proof. □

As the OTC stage progresses, the value of surplus declines while the cost of deficit rises. The direction of rate movement depends critically on market imbalance: rates decline in deficit markets as surplus becomes scarce, and rise in surplus markets as deficits become scarce.

A.8 Proof of Corollary 5

Proof. From Lemma 4.2, market tightness satisfies

$$\dot{\theta}_\tau = -\lambda\theta_\tau[\gamma(\theta_\tau^{-1}) - \gamma(\theta_\tau)].$$

Assume time τ has elapsed. Consider the solution of market tightness from time τ to τ' with initial condition θ_τ. Define the rescaled time $s = \frac{t-\tau}{1-\tau}$ for $t \in [\tau, 1]$, so that $s \in [0, 1]$. Then, $dt = (1-\tau)ds$.

By change of variables, the ODE for tightness becomes

$$\frac{d\theta}{ds} = (1-\tau)\frac{d\theta}{dt} = -(1-\tau)\lambda\theta[\gamma(\theta^{-1}) - \gamma(\theta)].$$

This is the same ODE, but with efficiency parameter $\lambda(1-\tau)$ and initial condition $\theta(0) = \theta_\tau$. At rescaled time $s' = \frac{\tau'-\tau}{1-\tau}$, we have

$$\theta(\tau', \theta_0, \lambda) = \theta\left(\frac{\tau'-\tau}{1-\tau}, \theta(\tau, \theta_0, \lambda), \lambda(1-\tau)\right).$$

For $\Psi^\pm$ the property follows from the market clearing condition $\bar{\theta} = \frac{1-\Psi^-_\tau}{1-\Psi^+_\tau}\theta_\tau$ and (A.12). For $\chi^\pm$ the property follows from Proposition 3. All of these objects can be expressed as a function of θ_t. Thus, the property is established immediately.

□

A.9 Proof of Corollary 6

Proof. **Part 1: Monotonicity in η.** We establish that χ^+, χ^-, and $\overline{r}^f$ are all decreasing in η. From the general formulas

$$\chi^+ = (r^w - r^m)\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{1-\eta}}{\bar{\theta} - 1}, \quad \chi^- = (r^w - r^m)\frac{\bar{\theta} - \bar{\theta}^\eta \theta_\tau^{-\eta}}{\bar{\theta} - 1},$$

we establish the result by taking derivatives with respect to η (holding θ_τ and $\bar{\theta}$ fixed)

$$\frac{\partial \chi^+}{\partial \eta} = -\frac{(r^w - r^m)}{\bar{\theta} - 1} \cdot \bar{\theta}^\eta \theta_\tau^{1-\eta} \ln\left(\frac{\bar{\theta}}{\theta_\tau}\right), \quad \frac{\partial \chi^-}{\partial \eta} = -\frac{(r^w - r^m)}{\bar{\theta} - 1} \cdot \bar{\theta}^\eta \theta_\tau^{-\eta} \ln\left(\frac{\bar{\theta}}{\theta_\tau}\right).$$

We establish that $\ln\left(\frac{\bar{\theta}}{\theta_\tau}\right)$, and $(\bar{\theta} - 1)$ always have the same sign: If (i) $\theta_0 > 1$: Market tightness increases over time, so $\bar{\theta} > \theta_\tau > 1$. Thus $\ln\left(\frac{\bar{\theta}}{\theta_\tau}\right) > 0$ and $(\bar{\theta} - 1) > 0$. Otherwise, if (ii) $\theta_0 < 1$: Market tightness decreases over time, so $\bar{\theta} < \theta_\tau < 1$. Thus, $\ln\left(\frac{\bar{\theta}}{\theta_\tau}\right) < 0$ and $(\bar{\theta} - 1) < 0$. At $\theta_0 = 1$: $\bar{\theta} = \theta_\tau = 1$, so both expressions equal zero.[40] Therefore, $\frac{\ln\left(\frac{\bar{\theta}}{\theta_\tau}\right)}{(\bar{\theta}-1)} > 0$ whenever $\theta_0 \neq 1$, which implies

$$\frac{\partial \chi^+}{\partial \eta} < 0, \quad \frac{\partial \chi^-}{\partial \eta} < 0.$$

The negotiated rate at time τ is: $r_\tau^f = r^m + (1-\eta)\chi_\tau^- + \eta\chi_\tau^+$. Differentiating with respect to η yields a negative value,

$$\frac{\partial r_\tau^f}{\partial \eta} = \underbrace{(\chi_\tau^+ - \chi_\tau^-)}_{<0} + \underbrace{(1-\eta)\frac{\partial \chi_\tau^-}{\partial \eta}}_{<0} + \underbrace{\eta\frac{\partial \chi_\tau^+}{\partial \eta}}_{<0} < 0.$$

Finally, we note that the average OTC rate $\overline{r}^f$ inherits this property since it is a weighted average of rates across all trading rounds, $\frac{\partial \overline{r}^f}{\partial \eta} < 0$.

Part 2: Limits. When $\eta = 1$

$$\chi_0^+ = (r^w - r^m)\int_0^1 0 \cdot \psi_y^+ \exp\left(-\int_y^1 \psi_x^- dx\right) dy = 0,$$
$$\chi_0^- = (r^w - r^m)\left[1 - \int_0^1 \psi_y^- \exp\left(-\int_y^1 \psi_x^- dx\right) dy\right] = (r^w - r^m)(1 - \Psi^-).$$

[40] We can evaluate the derivative at $\theta_0 = 1$ using L'Hospital's rule.

Therefore, $\bar{r}^f = r^m + 0 \cdot \chi^- + 1 \cdot \chi^+ = r^m$. When $\eta = 0$:

$$\chi_0^+ = (r^w - r^m) \int_0^1 \psi_y^+ \exp\left(-\int_y^1 \psi_x^+ dx\right) dy = (r^w - r^m)\Psi^+,$$
$$\chi_0^- = (r^w - r^m)\left[1 - \int_0^1 0 \cdot \psi_y^- \exp\left(-\int_y^1 \psi_x^+ dx\right) dy\right] = r^w - r^m.$$

Therefore, $\bar{r}^f = r^m + 1 \cdot \chi^- + 0 \cdot \chi^+ = r^m + (r^w - r^m) = r^w$. □

A.10 Proof of Corollary 7

Proof. **Part 1: Static limit ($\lambda \to 0$).** As $\lambda \to 0$, both matching intensities vanish: $\psi_\tau^\pm \to 0$. From Proposition 3

$$\chi^+ = (r^w - r^m) \int_0^1 (1-\eta)\psi_y^+ \exp\left(-\int_y^1 [\eta\psi_x^- + (1-\eta)\psi_x^+]dx\right) dy \to 0$$
$$\chi^- = (r^w - r^m)\left[1 - \int_0^1 \eta\psi_y^- \exp\left(-\int_y^1 [\eta\psi_x^- + (1-\eta)\psi_x^+]dx\right) dy\right] \to r^w - r^m.$$

Therefore, by the dilation property, all rates must equal: $\bar{r}^f \to r^m + (1-\eta)(r^w - r^m)$.

Part 2: Walrasian limit ($\lambda \to \infty$). First, consider a balance market where $\theta = 1$. As $\lambda \to \infty$, $\Sigma_0 \to 0$ as $\Psi^\pm \to 1$. From the balanced market formulas, it follows immediately that

$$\chi^\pm \to (r^w - r^m)(1-\eta), \quad \bar{r}^f = r^m + (1-\eta)(r^w - r^m).$$

When the market is unbalanced, we must be more careful. Start with $\theta > 1$. As $\lambda \to \infty$, the short (surplus) side matches instantly while the long (deficit) side faces rationing. This property follows because $\dot{\theta}_\tau = \lambda\gamma(\theta_\tau)(\theta_\tau - 1) > 0$, implying that tightness explodes: $\theta_\tau \to \infty$. Thus, surplus traders match instantly: $\Psi^+ \to 1$ whereas $\Psi^- = S^+/S^- \to \theta_0^{-1}$.

Hence, using the consistency property (Proposition 1), we have,

$$\chi_0^+ \to \bar{r}^f - r^m, \quad \chi_0^- \to (\bar{r}^f - r^m)\theta_0^{-1} + (r^w - r^m)(1 - \theta_0^{-1}).$$

Recall from Proposition 3 that the bargained rate at $\tau = 0$ satisfies,

$$r_0 - r^m = (1-\eta)\chi_0^- + \eta\chi_0^+.$$

Thus, as efficiency increases to infinity,

$$r_0^f - r^m \to (1-\eta)((\bar{r}^f - r^m)\theta_0^{-1} + (r^w - r^m)(1-\theta_0^{-1})) + \eta(\bar{r}^f - r^m).$$

We further know that as $\lambda \to \infty$, trades occur faster, i.e., matching intensities explode. Hence, the distribution of trade concentrates toward the earliest rounds, and thus the volume-weighted average rate converges to the first-trading round rate, $\bar{r}^f \to r_0^f$. Therefore, the limit above requires,

$$(r_0^f - r^m)(1-\eta)(1-\theta_0^{-1}) \to (r^w - r^m)(1-\eta)(1-\theta_0^{-1}) \Leftrightarrow r_0^f \to r^w.$$

By the dilation property, $r_\tau^f \to r^w, \forall \tau$. Thus, we conclude that when $\theta > 1$, as $\lambda \to \infty$, we have $\chi^\pm \to r^w - r^m$. When $\theta < 1$, by the symmetry property (Proposition 14), $\chi^\pm = 0, \bar{r}^f = r^m$. □

A.11 Proof of Corollary 8

Proof. The proof proceeds in two stages: forward induction to establish monotonicity of market tightness, followed by backward induction to establish monotonicity of outside options and rates.

Part 1: Forward induction on θ_n. Assume $\frac{\partial \theta_k}{\partial \theta_0} > 0$ for all $k \leq n$. From the market tightness recursion,

$$\theta_{n+1} = \theta_n \frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+},$$

where $\psi_{n+1}^+ = \lambda(N)G(1,\theta_n)$ and $\psi_{n+1}^- = \lambda(N)G(\theta_n^{-1},1)$. Since G is increasing in each argument,

$$\frac{\partial \psi_{n+1}^+}{\partial \theta_n} = \lambda(N)G_2(1,\theta_n) > 0, \quad \frac{\partial \psi_{n+1}^-}{\partial \theta_n} = \lambda(N)G_1(\theta_n^{-1},1)\cdot(-\theta_n^{-2}) < 0.$$

Taking the derivative of θ_{n+1} with respect to θ_0,

$$\frac{\partial \theta_{n+1}}{\partial \theta_0} = \frac{\partial \theta_n}{\partial \theta_0}\cdot\frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+} + \theta_n \frac{\partial}{\partial \theta_0}\left[\frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+}\right].$$

For $\theta_0 > 1$, we have $\theta_n > 1$ for all n. The first term is positive by the inductive hypothesis. The second term is also positive because as θ_n increases, the numerator $(1-\psi_{n+1}^-)$ increases (since $\frac{\partial \psi_{n+1}^-}{\partial \theta_n} < 0$) and the denominator $(1-\psi_{n+1}^+)$ decreases (since $\frac{\partial \psi_{n+1}^+}{\partial \theta_n} > 0$). Therefore, $\frac{\partial \theta_{n+1}}{\partial \theta_0} > 0$, completing the induction. This establishes Part 1 and implies,

$$\frac{\partial \psi_n^+}{\partial \theta_0} > 0 \quad \text{and} \quad \frac{\partial \psi_n^-}{\partial \theta_0} < 0 \quad \text{for all } n.$$

Part 2: Backward induction on outside options and rates. The terminal values are constants,

$$\chi_N^+ = 0 \implies \frac{\partial \chi_N^+}{\partial \theta_0} = 0, \quad \chi_N^- = r^w - r^m \implies \frac{\partial \chi_N^-}{\partial \theta_0} = 0.$$

We proceed by backward induction. Assume for round $n+1$ that $\frac{\partial \chi_{n+1}^+}{\partial \theta_0} \geq 0$, $\frac{\partial \chi_{n+1}^-}{\partial \theta_0} \geq 0$, and $\frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$. From the outside option recursions we obtain

$$\chi_n^+ = \psi_{n+1}^+(r_{n+1}^f - r^m) + (1 - \psi_{n+1}^+)\chi_{n+1}^+,$$
$$\chi_n^- = \psi_{n+1}^-(r_{n+1}^f - r^m) + (1 - \psi_{n+1}^-)\chi_{n+1}^-.$$

Analysis for χ_n^+: Differentiating with respect to θ_0, we have

$$\frac{\partial \chi_n^+}{\partial \theta_0} = \frac{\partial \psi_{n+1}^+}{\partial \theta_0}(r_{n+1}^f - r^m - \chi_{n+1}^+) + \psi_{n+1}^+ \frac{\partial r_{n+1}^f}{\partial \theta_0} + (1 - \psi_{n+1}^+)\frac{\partial \chi_{n+1}^+}{\partial \theta_0}.$$

From the bargaining solution, $r_{n+1}^f = r^m + \eta\chi_{n+1}^+ + (1-\eta)\chi_{n+1}^-$, which implies:

$$r_{n+1}^f - r^m - \chi_{n+1}^+ = (1-\eta)(\chi_{n+1}^- - \chi_{n+1}^+) \equiv (1-\eta)\Sigma_{n+1} > 0,$$

where $\Sigma_{n+1} > 0$ is the total surplus. Examining each term:

1. $\frac{\partial \psi_{n+1}^+}{\partial \theta_0} \cdot (1-\eta)\Sigma_{n+1} = (+)(+) > 0$,
2. $\psi_{n+1}^+ \cdot \frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$ (by inductive hypothesis),
3. $(1 - \psi_{n+1}^+) \cdot \frac{\partial \chi_{n+1}^+}{\partial \theta_0} \geq 0$ (by inductive hypothesis).

Therefore, $\frac{\partial \chi_n^+}{\partial \theta_0} > 0$. Similarly, differentiating, we obtain

$$\frac{\partial \chi_n^-}{\partial \theta_0} = \frac{\partial \psi_{n+1}^-}{\partial \theta_0}(r_{n+1}^f - r^m - \chi_{n+1}^-) + \psi_{n+1}^- \frac{\partial r_{n+1}^f}{\partial \theta_0} + (1 - \psi_{n+1}^-)\frac{\partial \chi_{n+1}^-}{\partial \theta_0}.$$

Now, $r_{n+1}^f - r^m - \chi_{n+1}^- = \eta(\chi_{n+1}^+ - \chi_{n+1}^-) = -\eta\Sigma_{n+1} < 0$. Examining the terms:

1. $\frac{\partial \psi_{n+1}^-}{\partial \theta_0} \cdot (-\eta\Sigma_{n+1}) = (-) \cdot (-) > 0$,
2. $\psi_{n+1}^- \cdot \frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$ (by inductive hypothesis),
3. $(1 - \psi_{n+1}^-) \cdot \frac{\partial \chi_{n+1}^-}{\partial \theta_0} \geq 0$ (by inductive hypothesis).

Therefore, $\frac{\partial \chi_n^-}{\partial \theta_0} > 0$.

Since $r_n^f = r^m + \eta\chi_n^+ + (1-\eta)\chi_n^-$:

$$\frac{\partial r_n^f}{\partial \theta_0} = \eta \frac{\partial \chi_n^+}{\partial \theta_0} + (1-\eta)\frac{\partial \chi_n^-}{\partial \theta_0} > 0.$$

This completes the backward induction for discrete rounds. □

Corollary 12 (Continuous-Time Limit)**.** *The monotonicity properties extend to the continuous-time limit. For $\theta_0 > 1$ and all $\tau \in (0,1)$:*

$$\frac{\partial \chi_\tau^+}{\partial \theta_0} > 0, \quad \frac{\partial \chi_\tau^-}{\partial \theta_0} > 0, \quad \frac{\partial r_\tau^f}{\partial \theta_0} > 0.$$

Proof. As $N \to \infty$, the discrete-time equilibrium converges uniformly to the continuous-time equilibrium on $[0,1]$. Since strict inequalities are preserved under uniform convergence, the monotonicity properties carry over to the continuous-time limit. □

A.12 Proof of Proposition 4

Proof. We first prove the result for $\theta \to 0$ and then use symmetry to show the opposite limit. As $\theta \to 0$,

$$\frac{\dot{\theta}}{\theta} = -\lambda[\gamma(\theta^{-1}) - \gamma(\theta)] \approx -\lambda\gamma(\theta^{-1}) \to -\lambda\bar{\gamma},$$

where we used the no-disposal property of the matching function to establish that $\gamma(\theta) \to 0$ as $\theta \to 0$. If $\bar{\gamma} < \infty$, matching intensities are finite, so trade is active throughout all rounds. Moreover, θ_τ approximately satisfies exponential decay, $\theta_\tau \sim \theta_0 e^{-\lambda\bar{\gamma}\tau}$, so that the terminal tightness satisfies $\bar{\theta}/\theta_0 \to e^{-\lambda\bar{\gamma}}$.

For surplus traders, we have

$$\Psi^+ = 1 - \exp\left(-\int_0^1 \lambda\gamma(\theta_\tau)d\tau\right) \to 0$$

since $\gamma(\theta) \to 0$ as $\theta \to 0$.

For deficit traders, if $\bar{\gamma} < \infty$,

$$\Psi^- = 1 - \exp\left(-\int_0^1 \lambda\gamma(\theta_\tau^{-1})d\tau\right) \approx 1 - \exp(-\lambda\bar{\gamma}).$$

If $\bar{\gamma} = \infty$, then $\Psi^- \to 1$ as $\theta \to 0$. From Proposition 3, for finite $\bar{\gamma}$,

$$\chi^+ \to (r^w - r^m) \int_0^1 (1-\eta)\psi_y^+ \Sigma_y dy \to 0,$$

given that ψ_y^+ converges uniformly to zero as $\theta \to 0$ and the surplus is always finite. For χ^- we have that

$$\chi^- \to (r^w - r^m)\left[1 - \int_0^1 \eta\lambda\bar{\gamma}e^{-\eta\lambda\bar{\gamma}(1-y)}dy\right] = (r^w - r^m)e^{-\lambda\bar{\gamma}\eta}.$$

The first limit passes limits inside the integral and uses the continuity of $\exp(\cdot)$.

For $\bar{\gamma}$ infinite, we bound χ^- above by $(r^w - r^m)e^{-\lambda\gamma(\theta_0)\eta}$, noting that $\gamma(\theta_\tau)$ is increasing in time. Then apply the limit as $\theta_0 \to \infty$ to find that $\chi^- = 0$. Hence, the solution above remains valid.

For the average interbank rate, we use Proposition 3. Since $\bar{\theta}/\theta_0 \to e^{-\lambda\bar{\gamma}}$ as $\theta_0 \to 0$, we have

$$\phi = \frac{(\bar{\theta}/\theta_0)^\eta - 1}{\bar{\theta}/\theta_0 - 1} \to \frac{e^{-\lambda\bar{\gamma}\eta} - 1}{e^{-\lambda\bar{\gamma}} - 1}.$$

Hence,

$$\bar{r}^f - r^m = (1-\phi)(r^w - r^m) = (r^w - r^m) \cdot \frac{e^{(1-\eta)\lambda\bar{\gamma}} - 1}{e^{\lambda\bar{\gamma}} - 1}.$$

Finally, we obtain the corresponding formulas for $\theta \to \infty$ by symmetry. Using Proposition 14, this is case (i) with $\theta \leftrightarrow \theta^{-1}$, $\psi^+ \leftrightarrow \psi^-$, and $\eta \leftrightarrow 1-\eta$. Thus,

$$\Psi^+ = 1 - e^{-\lambda\bar{\gamma}}, \quad \Psi^- = 0, \quad \chi^+ = (r^w - r^m)(1 - e^{-(1-\eta)\lambda\bar{\gamma}}), \quad \chi^- = r^w - r^m.$$

For the average rate, since $\bar{\theta}/\theta_0 \to e^{\lambda\bar{\gamma}}$ as $\theta_0 \to \infty$,

$$\bar{r}^f - r^m = (r^w - r^m) \cdot \frac{1 - e^{-\lambda\bar{\gamma}(1-\eta)}}{1 - e^{-\lambda\bar{\gamma}}}.$$

□

A.13 Proof of Corollary 9

Proof. Write the ODE in Equation (12) as

$$\dot{\theta}_\tau = \lambda F(\theta_\tau), \qquad \theta_0 = \theta_0,$$

where $F(\theta) \equiv -\theta\big(\gamma(\theta^{-1}) - \gamma(\theta)\big)$. We perform a change of variables: $u \equiv \lambda\tau$. Then, using that $d\tau = \lambda du$, we have that $\vartheta(u)$ satisfies:

$$\vartheta'(u) = F(\vartheta(u)), \qquad \vartheta(0) = \theta_0.$$

By existence and uniqueness of solutions, $\theta(\tau;\theta_0,\lambda) = \vartheta(\lambda\tau)$ for all $\tau \in [0,1]$. Differentiating with respect to λ yields

$$\frac{\partial}{\partial\lambda}\theta(\tau;\theta_0,\lambda) = \tau\,\vartheta'(\lambda\tau) = \tau\,F(\theta(\tau;\theta_0,\lambda))\,.$$

Therefore, for $\tau \in (0,1]$, the sign of $\partial_\lambda\theta(\tau;\theta_0,\lambda)$ is the sign of $F(\theta)$ evaluated along the solution θ_t.

Recall that, $\gamma(\theta) = \theta\,\gamma(\theta^{-1})$. Hence

$$F(\theta) = -\theta\big(\gamma(\theta^{-1}) - \gamma(\theta)\big) = -\theta\big(1-\theta\big)\gamma(\theta^{-1}) = \theta(\theta-1)\gamma(\theta^{-1}).$$

We thus obtain that $\operatorname{sign}(F(\theta)) = \operatorname{sign}(\theta - 1)$.

Corollary 2 then implies that if $\theta_0 > 1$ then $\theta(\tau;\theta_0,\lambda) > 1$ for all $\tau \in (0,1]$, and similarly if $\theta_0 < 1$ then $\theta(\tau;\theta_0,\lambda) < 1$ for all $\tau \in (0,1]$. Thus $\operatorname{sign}(F(\theta(\tau))) = \operatorname{sign}(\theta_0 - 1)$ for all $\tau \in (0,1]$, completing the proof. □

A.14 Proof of Proposition 5

Proof. Recall that

$$\chi^+ = (r^w - r^m)\frac{\bar\theta - \bar\theta^\eta\theta^{1-\eta}}{\bar\theta - 1}, \quad \chi^- = (r^w - r^m)\frac{\bar\theta - \bar\theta^\eta\theta^{-\eta}}{\bar\theta - 1},$$

where $\bar\theta = \bar\theta(\theta,\lambda)$ is the terminal market tightness given in Table 1. Define

$$h^+(\bar\theta) \equiv \frac{\bar\theta - \bar\theta^\eta\theta^{1-\eta}}{\bar\theta - 1}, \quad h^-(\bar\theta) \equiv \frac{\bar\theta - \bar\theta^\eta\theta^{-\eta}}{\bar\theta - 1}.$$

Corollary 9 allows as to establish the monotonicity of $\bar\theta$ in λ which is key for the proof.

Step 1: Derivatives with respect to $\bar{\theta}$. For $h^+(\bar{\theta})$, using the quotient rule

$$\begin{aligned} h^+_{\bar{\theta}}(\bar{\theta}) &= \frac{[1-\eta\bar{\theta}^{\eta-1}\theta^{1-\eta}](\bar{\theta}-1)-[\bar{\theta}-\bar{\theta}^{\eta}\theta^{1-\eta}]}{(\bar{\theta}-1)^2} \\ &= \frac{\eta\bar{\theta}^{\eta-1}\theta^{1-\eta}+(1-\eta)\bar{\theta}^{\eta}\theta^{1-\eta}-1}{(\bar{\theta}-1)^2} \\ &= \frac{\bar{\theta}^{\eta}\theta^{1-\eta}[(1-\eta)\bar{\theta}+\eta]-\bar{\theta}}{(\bar{\theta}-1)^2\cdot\bar{\theta}}. \end{aligned}$$

Similarly, for $h^-(\bar{\theta})$,

$$h^-_{\bar{\theta}}(\bar{\theta}) = \frac{\bar{\theta}^{\eta}\theta^{-\eta}[(1-\eta)\bar{\theta}+\eta]-\bar{\theta}}{(\bar{\theta}-1)^2\cdot\bar{\theta}}.$$

Since $(\bar{\theta}-1)^2>0$ and $\bar{\theta}>0$, the sign of $h^{\pm}_{\bar{\theta}}$ is determined by

$$\text{sign}(h^{\pm}_{\bar{\theta}}) = \text{sign}\left(\bar{\theta}^{\eta}\theta^{\frac{1}{2}\pm\frac{1}{2}-\eta}[(1-\eta)\bar{\theta}+\eta]-\bar{\theta}\right). \tag{A.13}$$

Step 2: Case $\theta > 1$. We test the monotonicity of χ^+ by evaluating condition (A.13). For that, consider

$$D(\bar{\theta};\theta) = \theta^{1-\eta}[(1-\eta)\bar{\theta}+\eta]-\bar{\theta}^{1-\eta}.$$

By Corollary 2, $\bar{\theta}\geq\theta$, Evaluated at $\bar{\theta}=\theta$,

$$\begin{aligned} D(\bar{\theta}) &= \theta^{1-\eta}[(1-\eta)\theta+\eta]-\theta^{1-\eta} \\ &= \theta^{1-\eta}[\theta-\eta(\theta-1)]-\theta^{1-\eta} \\ &= \theta^{2-\eta}-\eta\theta^{1-\eta}(\theta-1)-\theta^{1-\eta} \\ &> (\theta-1)\theta^{1-\eta}>0, \end{aligned}$$

where the inequalities follow from $\theta>1$—this implies $\theta^{2-\eta}>\theta^{1-\eta}$ and $\eta\theta^{1-\eta}(\theta-1)>0$.

Taking the derivative

$$D'(\bar{\theta}) = (1-\eta)\theta^{1-\eta}-(1-\eta)\bar{\theta}^{-\eta} = (1-\eta)(\theta^{1-\eta}-\bar{\theta}^{-\eta}).$$

The term in parentheses is positive since $\theta,\bar{\theta}>1$. Since $D(\theta)>0$ and $D'(\bar{\theta})>0$ for $\bar{\theta}>\theta$, we have $D(\bar{\theta})>0$ for all $\bar{\theta}\geq\theta$. Therefore $h^+_{\bar{\theta}}>0$ for all $\bar{\theta}\geq\theta$, and

$$\frac{\partial\chi^+}{\partial\lambda} = (r^w-r^m)h^+_{\bar{\theta}}\cdot\frac{\partial\bar{\theta}}{\partial\lambda}>0.$$

Next, we test the monotonicity of χ^- by evaluating condition (A.13). For that, consider

$$F(\bar{\theta};\theta) = \theta^{-\eta}[(1-\eta)\bar{\theta}+\eta] - \bar{\theta}^{1-\eta}.$$

Evaluated at $\bar{\theta} = \theta$,

$$\begin{aligned} F(\bar{\theta}) = \theta^{-\eta}[(1-\eta)\theta+\eta] - \theta^{1-\eta} &= \theta^{-\eta}[\theta - \eta(\theta-1)] - \theta^{1-\eta} \\ &= \theta^{1-\eta} - \eta\theta^{-\eta}(\theta-1) - \theta^{1-\eta} < 0. \end{aligned}$$

Hence, the sign is negative at $\bar{\theta} = \theta$.

As $\bar{\theta} \to \infty$, we have the following limit:

$$\lim_{\bar{\theta}\to\infty} F(\bar{\theta}) = \theta^{1-\eta}\eta + \lim_{\bar{\theta}\to\infty} \bar{\theta}[(1-\eta)\theta^{-\eta} - \bar{\theta}^{-\eta}] \to \infty > 0,$$

By continuity of $\bar{\theta}$ in λ, $F(\bar{\theta})$ changes sign on $[\theta, \infty)$. Hence, χ^- is non-monotonic in λ as $\theta > 1$.

We further verify that the function changes sign only once. For that, observe that:

$$D'(\bar{\theta}) = (1-\eta)\theta^{-\eta} - (1-\eta)\bar{\theta}^{-\eta} = (1-\eta)(\theta^{-\eta} - \bar{\theta}^{-\eta}) > 0,$$

where the inequality follows from $\bar{\theta} > \theta > 1$. Thus, χ^-_λ changes sing only once.

Finally, we check the monotonicity of $\bar{r}^f$. The average interest rate can be written as

$$\bar{r}^f = \phi(\theta) r^m + (1-\phi(\theta)) r^w,$$

where the endogenous bargaining power is given by

$$\phi(\theta) = \begin{cases} 1 - \dfrac{\bar{\theta} - \theta^{1-\eta}\bar{\theta}^{\eta}}{\bar{\theta} - \theta} & \text{if } \theta > 1, \\ \eta & \text{if } \theta = 1, \\ 1 - \dfrac{\theta^{-1} - \theta^{-\eta}\bar{\theta}^{\eta-1}}{\bar{\theta}^{-1} - \theta^{-1}} & \text{if } \theta < 1. \end{cases}$$

For $\theta > 1$, note that $\phi(\theta) = 1 - h^+_{\bar{\theta}}(\bar{\theta})$. Since we have shown that $h^+_{\bar{\theta}} > 0$,

$$\frac{\partial \phi}{\partial \lambda} = -\frac{\partial h^+}{\partial \bar{\theta}} \cdot \frac{\partial \bar{\theta}}{\partial \lambda} < 0.$$

Therefore,

$$\frac{\partial \bar{r}^f}{\partial \lambda} = \frac{\partial \phi}{\partial \lambda}(r^m - r^w) = -\frac{\partial \phi}{\partial \lambda}(r^w - r^m) > 0.$$

Thus $\bar{r}^f$ is increasing in λ when $\theta > 1$.

Step 3: Case $\theta < 1$. The results for $\theta < 1$ follow by symmetry (Proposition 14).

Step 4: Case $\theta = 1$. When $\theta = 1$, $\bar{\theta} = 1$ for all λ, so $\bar{r}^f = r^m + (1-\eta)(r^w - r^m)$ is constant. Using the explicit formulas,

$$\chi^+ = \Psi^+(\bar{r}^f - r^m), \quad \chi^- = (r^w - r^m) - \Psi^-(r^w - \bar{r}^f),$$

where $\Psi^\pm = 1 - e^{-\lambda}$. Therefore, we obtain

$$\frac{\partial \chi^+}{\partial \lambda} = (1-\eta)(r^w - r^m)e^{-\lambda} > 0, \quad \frac{\partial \chi^-}{\partial \lambda} = -\eta(r^w - r^m)e^{-\lambda} < 0.$$

□

A.15 Proof of Corollary 10

Proof. When $\theta > 1$, the OTC rate is decreasing over time (Corollary 4), so $Q = r_0^f - r_1^f$. When $\theta < 1$, the rate is increasing, so $Q = r_1^f - r_0^f$. Since the terminal rate r_1^f is constant, all variation in Q comes from the initial rate r_0^f. From Corollary 8, $\bar{r}^f$ is increasing in both θ and λ when $\theta > 1$, and decreasing in both when $\theta < 1$. Hence, considering the time dilation property, we conclude that for $\theta > 1$, both $\frac{\partial Q}{\partial \theta} > 0$ and $\frac{\partial Q}{\partial \lambda} > 0$. For $\theta < 1$, the sign of Q reverses, so $\frac{\partial Q}{\partial \theta} = -\frac{\partial \bar{r}^f}{\partial \theta} < 0$ while $\frac{\partial Q}{\partial \lambda} = -\frac{\partial \bar{r}^f}{\partial \lambda} > 0$. When $\theta = 1$, the rate is constant over time, so $Q = 0$ and both derivatives vanish. □

A.16 Proof of Corollary 11

Proof. From the matching consistency condition, the deficit-side matching probability is:

$$\Psi^- = \frac{\bar{\theta} - \theta_0}{(\bar{\theta} - 1)\theta_0},$$

where $\bar{\theta} \equiv \theta_1$ denotes the terminal market tightness. Differentiating with respect to $\bar{\theta}$:

$$\frac{\partial \Psi^-}{\partial \bar{\theta}} = \frac{(\bar{\theta} - 1) - (\bar{\theta} - \theta_0)}{(\bar{\theta} - 1)^2 \theta_0} = \frac{\theta_0 - 1}{(\bar{\theta} - 1)^2 \theta_0}.$$

Thus, $\partial \Psi^- / \partial \bar{\theta} > 0$ if $\theta_0 > 1$ and $\partial \Psi^- / \partial \bar{\theta} < 0$ if $\theta_0 < 1$.

By Corollary 9, the terminal tightness satisfies $\text{sign}(\partial \bar{\theta} / \partial \lambda) = \text{sign}(\theta_0 - 1)$. Applying the chain rule:

$$\frac{\partial \Psi^-}{\partial \lambda} = \frac{\partial \Psi^-}{\partial \bar{\theta}} \cdot \frac{\partial \bar{\theta}}{\partial \lambda}.$$

When $\theta_0 > 1$, both terms are positive, so $\partial \Psi^- / \partial \lambda > 0$. When $\theta_0 < 1$, both terms are negative, so $\partial \Psi^- / \partial \lambda > 0$. Hence, Ψ^- is increasing in λ for all $\theta_0 \neq 1$.

Since $I = (1 - \Psi^-)/\Psi^-$ is decreasing in Ψ^-, it follows that I is decreasing in λ. □

B. Analytic Solutions: Calculations for Table 1

Proof. The proof proceeds in two stages: forward induction to establish monotonicity of market tightness, followed by backward induction to establish monotonicity of outside options and rates.

Part 1: Forward induction on θ_n. Assume $\frac{\partial \theta_k}{\partial \theta_0} > 0$ for all $k \leq n$. From the market tightness recursion,

$$\theta_{n+1} = \theta_n \frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+},$$

where $\psi_{n+1}^+ = \lambda(N)G(1,\theta_n)$ and $\psi_{n+1}^- = \lambda(N)G(\theta_n^{-1},1)$. Since G is increasing in each argument,

$$\frac{\partial \psi_{n+1}^+}{\partial \theta_n} = \lambda(N)G_2(1,\theta_n) > 0, \quad \frac{\partial \psi_{n+1}^-}{\partial \theta_n} = \lambda(N)G_1(\theta_n^{-1},1)\cdot(-\theta_n^{-2}) < 0.$$

Taking the derivative of θ_{n+1} with respect to θ_0,

$$\frac{\partial \theta_{n+1}}{\partial \theta_0} = \frac{\partial \theta_n}{\partial \theta_0}\cdot\frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+} + \theta_n \frac{\partial}{\partial \theta_0}\left[\frac{1-\psi_{n+1}^-}{1-\psi_{n+1}^+}\right].$$

For $\theta_0 > 1$, we have $\theta_n > 1$ for all n. The first term is positive by the inductive hypothesis. The second term is also positive because as θ_n increases, the numerator $(1-\psi_{n+1}^-)$ increases (since $\frac{\partial \psi_{n+1}^-}{\partial \theta_n} < 0$) and the denominator $(1-\psi_{n+1}^+)$ decreases (since $\frac{\partial \psi_{n+1}^+}{\partial \theta_n} > 0$). Therefore, $\frac{\partial \theta_{n+1}}{\partial \theta_0} > 0$, completing the induction. This establishes Part 1 and implies,

$$\frac{\partial \psi_n^+}{\partial \theta_0} > 0 \quad \text{and} \quad \frac{\partial \psi_n^-}{\partial \theta_0} < 0 \quad \text{for all } n.$$

Part 2: Backward induction on outside options and rates. The terminal values are constants,

$$\chi_N^+ = 0 \implies \frac{\partial \chi_N^+}{\partial \theta_0} = 0, \quad \chi_N^- = r^w - r^m \implies \frac{\partial \chi_N^-}{\partial \theta_0} = 0.$$

We proceed by backward induction. Assume for round $n+1$ that $\frac{\partial \chi_{n+1}^+}{\partial \theta_0} \geq 0$, $\frac{\partial \chi_{n+1}^-}{\partial \theta_0} \geq 0$, and $\frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$. From the outside option recursions we obtain

$$\chi_n^+ = \psi_{n+1}^+(r_{n+1}^f - r^m) + (1-\psi_{n+1}^+)\chi_{n+1}^+,$$
$$\chi_n^- = \psi_{n+1}^-(r_{n+1}^f - r^m) + (1-\psi_{n+1}^-)\chi_{n+1}^-.$$

Analysis for χ_n^+: Differentiating with respect to θ_0, we have

$$\frac{\partial \chi_n^+}{\partial \theta_0} = \frac{\partial \psi_{n+1}^+}{\partial \theta_0}(r_{n+1}^f - r^m - \chi_{n+1}^+) + \psi_{n+1}^+ \frac{\partial r_{n+1}^f}{\partial \theta_0} + (1 - \psi_{n+1}^+)\frac{\partial \chi_{n+1}^+}{\partial \theta_0}.$$

From the bargaining solution, $r_{n+1}^f = r^m + \eta\chi_{n+1}^+ + (1-\eta)\chi_{n+1}^-$, which implies:

$$r_{n+1}^f - r^m - \chi_{n+1}^+ = (1-\eta)(\chi_{n+1}^- - \chi_{n+1}^+) \equiv (1-\eta)\Sigma_{n+1} > 0,$$

where $\Sigma_{n+1} > 0$ is the total surplus. Examining each term:

1. $\frac{\partial \psi_{n+1}^+}{\partial \theta_0} \cdot (1-\eta)\Sigma_{n+1} = (+)(+) > 0$,
2. $\psi_{n+1}^+ \cdot \frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$ (by inductive hypothesis),
3. $(1 - \psi_{n+1}^+) \cdot \frac{\partial \chi_{n+1}^+}{\partial \theta_0} \geq 0$ (by inductive hypothesis).

Therefore, $\frac{\partial \chi_n^+}{\partial \theta_0} > 0$. Similarly, differentiating, we obtain

$$\frac{\partial \chi_n^-}{\partial \theta_0} = \frac{\partial \psi_{n+1}^-}{\partial \theta_0}(r_{n+1}^f - r^m - \chi_{n+1}^-) + \psi_{n+1}^- \frac{\partial r_{n+1}^f}{\partial \theta_0} + (1 - \psi_{n+1}^-)\frac{\partial \chi_{n+1}^-}{\partial \theta_0}.$$

Now, $r_{n+1}^f - r^m - \chi_{n+1}^- = \eta(\chi_{n+1}^+ - \chi_{n+1}^-) = -\eta\Sigma_{n+1} < 0$. Examining the terms:

1. $\frac{\partial \psi_{n+1}^-}{\partial \theta_0} \cdot (-\eta\Sigma_{n+1}) = (-) \cdot (-) > 0$,
2. $\psi_{n+1}^- \cdot \frac{\partial r_{n+1}^f}{\partial \theta_0} \geq 0$ (by inductive hypothesis),
3. $(1 - \psi_{n+1}^-) \cdot \frac{\partial \chi_{n+1}^-}{\partial \theta_0} \geq 0$ (by inductive hypothesis).

Therefore, $\frac{\partial \chi_n^-}{\partial \theta_0} > 0$.

Since $r_n^f = r^m + \eta\chi_n^+ + (1-\eta)\chi_n^-$:

$$\frac{\partial r_n^f}{\partial \theta_0} = \eta\frac{\partial \chi_n^+}{\partial \theta_0} + (1-\eta)\frac{\partial \chi_n^-}{\partial \theta_0} > 0.$$

This completes the backward induction for discrete rounds. □

Corollary 13 (Continuous-Time Limit)**.** *The monotonicity properties extend to the continuous-time limit. For $\theta_0 > 1$ and all $\tau \in (0,1)$:*

$$\frac{\partial \chi_\tau^+}{\partial \theta_0} > 0, \quad \frac{\partial \chi_\tau^-}{\partial \theta_0} > 0, \quad \frac{\partial r_\tau^f}{\partial \theta_0} > 0.$$

Proof. As $N \to \infty$, the discrete-time equilibrium converges uniformly to the continuous-time equilibrium on $[0, 1]$. Since strict inequalities are preserved under uniform convergence, the monotonicity properties carry over to the continuous-time limit. □

C. Additional Results

C.1 Symmetry of Convenience yield

Symmetry of the matching function carries over the matching intensities OTC rate yield coefficients in the sense that the difference between r^w and χ^- is equal to the difference between r^w and χ^-

Corollary 14 (Symmetry). *The OTC market satisfies:* $(i)\, r^f(\tau,\theta,\eta,\lambda) = (r^w - r^m) - r^f(\tau\theta^{-1}, 1-\eta, \lambda)$;
$(ii)\, \chi^-(\tau,\theta,\eta,\lambda) = (r^w - r^m) - \chi^+(\tau,\theta^{-1},1-\eta,\lambda)$;
$(iii)\, \chi^+(\tau,\theta,\eta,\lambda) = (r^w - r^m) - \chi^-(\tau,\theta^{-1},1-\eta,\lambda)$.

This symmetry states that if we can reverse the market tightness and exchange the bargaining power of surplus and deficit sides, the surplus function is the same. For the expected cost of being in deficit, the benefit of a surplus, and the OTC rate, the same symmetry holds relative to $r^w - r^m$. A similar pattern holds for the average OTC rate. This property reveals that any asymmetry in outcomes follows from asymmetries in assumed bargaining powers or the tightness. This symmetry property implies that we can solve the model for $\theta < 1$ and obtain solutions for $\theta > 1$, immediately. Furthermore, as we will see, many properties change exactly at $\theta = 1$, due to this property.

Proof. **Step 1: Symmetry of market tightness.** From the ODE in Lemma 4.2:

$$\dot{\theta}_\tau = -\lambda\theta_\tau[\gamma(\theta_\tau^{-1}) - \gamma(\theta_\tau)]$$

Define, $\phi_\tau \equiv \theta_\tau^{-1}$, we have $\dot{\phi}_\tau = -\theta_\tau^{-2}\dot{\theta}_\tau$. Substituting,

$$\dot{\phi}_\tau = -\phi_\tau^2 \cdot (-\lambda\phi_\tau^{-1})[\gamma(\phi_\tau) - \gamma(\phi_\tau^{-1})] = -\lambda\phi_\tau[\gamma(\phi_\tau^{-1}) - \gamma(\phi_\tau)].$$

This is the same ODE as for θ. Hence, if θ_τ solves it with initial condition θ_0, then θ_τ^{-1} solves it with initial condition θ_0^{-1}.

Step 2: Symmetry of matching intensities. By definition,

$$\psi^+(\theta) = \lambda\gamma(\theta) = \lambda\theta\gamma(\theta^{-1}) = \lambda\theta\gamma(\theta^{-1}) = \psi^-(\theta^{-1}).$$

Step 3: Symmetry of the spread. From Proposition 3:

$$\Sigma_\tau(\theta,\eta) = (r^w - r^m)\exp\left(-\int_\tau^1 [\eta\psi_s^-(\theta_s) + (1-\eta)\psi_s^+(\theta_s)]ds\right).$$

Using $\psi^+(\theta) = \psi^-(\theta^{-1})$ and $\psi^-(\theta) = \psi^+(\theta^{-1})$:

$$\begin{aligned}\Sigma_\tau(\theta^{-1}, 1-\eta) &= (r^w - r^m)\exp\left(-\int_\tau^1 [(1-\eta)\psi_s^-(\theta_s^{-1}) + \eta\psi_s^+(\theta_s^{-1})]ds\right)\\ &= (r^w - r^m)\exp\left(-\int_\tau^1 [(1-\eta)\psi_s^+(\theta_s) + \eta\psi_s^-(\theta_s)]ds\right)\\ &= \Sigma_\tau(\theta, \eta).\end{aligned}$$

Step 4: Symmetry of convenience yields. From the expressions in Proposition 3:

$$\begin{aligned}\chi_\tau^+(\theta^{-1}, 1-\eta) &= (r^w - r^m)\int_\tau^1 \eta\psi_y^+(\theta_y^{-1})\Sigma_y(\theta^{-1}, 1-\eta)dy\\ &= (r^w - r^m)\int_\tau^1 \eta\psi_y^-(\theta_y)\Sigma_y(\theta, \eta)dy\\ &= (r^w - r^m) - \chi_\tau^-(\theta, \eta).\end{aligned}$$

Thus, $\chi_\tau^-(\theta^{-1}, 1-\eta) = (r^w - r^m) - \chi_\tau^+(\theta, \eta)$.

Step 5: Symmetry of the OTC rate. Using $r^f = r^m + (1-\eta)\chi^- + \eta\chi^+$ the symmetry follows from the symmetry of outside options. □

C.2 Derivation of investors' first-order condition, eq. (16)

After replacing the budget constraint (2) into the objective (15), we have the following objective:

$$\max_{\{a_i\},\, m}\left(\mathbb{E}_{X,\omega}\left[\left(R_m e + \sum_{i\in I}(R^i(X_t) - R^m)a_i + \chi_{t+1}(s(\{a_i\}_{i\in I}, m; \omega))\right)^{1-\gamma}\right]\right)^{\frac{1}{1-\gamma}}.$$

The first-order condition with respect to each asset a^i and to cash m are

$$a^i: \quad \mathbb{E}_{X,\omega}\Big[(R^e)^{-\gamma}\big(R^i(X_t) - R^m + \chi_s \tfrac{\partial s}{\partial a^i}\big)\Big] = 0,$$

$$m: \quad \mathbb{E}_{X,\omega}\Big[(R^e)^{-\gamma}\big(R^m - R^m + \chi_s \tfrac{\partial s}{\partial m}\big)\Big] = 0,$$

where R^e is the investor's total portfolio return and $\chi_s = \partial\chi/\partial s$ is the marginal convenience yield.

Combining these two first-order conditions:

$$\mathbb{E}_{X,\omega}\Big[(R^e)^{-\gamma}\big(R^i(X_t) - R^m + \chi_s(\tfrac{\partial s}{\partial a^i} - \tfrac{\partial s}{\partial m})\big)\Big] = 0.$$

Re-arranging gives

$$\mathbb{E}_X[R^i(X_t)] - R^m = \mathbb{E}_{X,\omega}\Big[\chi_s\big(\tfrac{\partial s}{\partial m} - \tfrac{\partial s}{\partial a^i}\big)\Big] - \frac{\mathbb{COV}_{X,\omega}\Big[(R^e)^{-\gamma},\, R^i(X_t) + \chi_s\big(\tfrac{\partial s}{\partial a^i} - \tfrac{\partial s}{\partial m}\big)\Big]}{\mathbb{E}_{X,\omega}[(R^e)^{-\gamma}]}.$$

Expand the covariance term and collecting like terms:

$$\begin{aligned}\mathbb{E}_X[R^i(X_t)] - R^m = \mathbb{E}_{X,\omega}\big[\chi_s\big(\tfrac{\partial s}{\partial m} - \tfrac{\partial s}{\partial a^i}\big)\big] \\ - \frac{\mathbb{COV}_{X,\omega}\Big[(R^e)^{-\gamma},\, \chi_s\big(\tfrac{\partial s}{\partial m} - \tfrac{\partial s}{\partial a^i}\big)\Big]}{\mathbb{E}_{X,\omega}[(R^e)^{-\gamma}]} - \frac{\mathbb{COV}_{X,\omega}\Big[(R^e)^{-\gamma}, R^i(X_t)\Big]}{\mathbb{E}_{X,\omega}[(R^e)^{-\gamma}]}.\end{aligned}$$

Rearranging this equation yields (16) in the text.

C.3 Derivation of planner's first-order condition, eq. (19)

The planner's optimal portfolio satisfies

$$a^i: \quad \mathbb{E}_\omega\left[(R^e)^{-\gamma}\Big(R^i - R^m + \chi_s\frac{\partial s}{\partial a^i} + \chi_\theta\frac{\partial \theta}{\partial a^i}\Big)\right] = 0,$$

$$m: \quad \mathbb{E}_\omega\left[(R^e)^{-\gamma}\Big(R^m - R^m + \chi_s\frac{\partial s}{\partial m} + \chi_\theta\frac{\partial \theta}{\partial m}\Big)\right] = 0,$$

where $\chi_s = \partial\chi/\partial s$ and $\chi_\theta = \partial\chi/\partial\theta$.

Combining these two:

$$\mathbb{E}_\omega\left[(R^e)^{-\gamma}\Big(R^i - R^m + \chi_s\Big(\tfrac{\partial s}{\partial a^i} - \tfrac{\partial s}{\partial m}\Big) + \chi_\theta\Big(\tfrac{\partial \theta}{\partial a^i} - \tfrac{\partial \theta}{\partial m}\Big)\Big)\right] = 0.$$

Re-arranging gives

$$R^i - R^m = \mathbb{E}_\omega\left[\chi_s\Big(\frac{\partial s}{\partial m} - \frac{\partial s}{\partial a^i}\Big) + \chi_\theta\Big(\frac{\partial \theta}{\partial m} - \frac{\partial \theta}{\partial a^i}\Big)\right] - \frac{\mathbb{COV}_\omega\Big[(R^e)^{-\gamma}, R^i + \chi_s\big(\tfrac{\partial s}{\partial a^i} - \tfrac{\partial s}{\partial m}\big) + \chi_\theta\big(\tfrac{\partial \theta}{\partial a^i} - \tfrac{\partial \theta}{\partial m}\big)\Big]}{\mathbb{E}_\omega[(R^e)^{-\gamma}]}.$$

Expanding the covariance term and collecting like terms yields

$$R^i - R^m = \mathbb{E}_\omega\left[\chi_s\left(\tfrac{\partial s}{\partial m} - \tfrac{\partial s}{\partial a^i}\right) + \chi_\theta\left(\tfrac{\partial \theta}{\partial m} - \tfrac{\partial \theta}{\partial a^i}\right)\right]$$

$$- \frac{\mathbb{COV}_\omega\left[(R^e)^{-\gamma}, \chi_s\left(\tfrac{\partial s}{\partial m} - \tfrac{\partial s}{\partial a^i}\right) + \chi_\theta\left(\tfrac{\partial \theta}{\partial m} - \tfrac{\partial \theta}{\partial a^i}\right)\right]}{\mathbb{E}_\omega[(R^e)^{-\gamma}]} - \frac{\mathbb{COV}_\omega\left[(R^e)^{-\gamma}, R^i\right]}{\mathbb{E}_\omega[(R^e)^{-\gamma}]}.$$

Rearranging and using the risk-adjusted probability measure Φ yields (19).