OTC Frictions and Price Formation in FX Markets*

Iñaki Aldasoro Mauricio Calani Damian Romero

February 2026

Preliminary and incomplete

Abstract

This paper develops a model of joint spot and forward exchange rate determination in a small open economy where spot foreign exchange rate (FX) and bonds trade competitively, while FX forwards are negotiated over-the-counter (OTC) under search frictions and bilateral bargaining. Clients (ranging from natural hedgers with exogenous FX exposures to global investors with speculative demand) choose forward notionals under mean–variance preferences. Dealers intermediate client trades, hedge inventory in spot and bond markets, and face convex balance-sheet costs that increase with market tightness (clients per dealer) and a financial stress state. The model generates an endogenous forward basis, predicts how tightness and stress reshape bargaining power, prices, and volumes, and delivers feedback from OTC trading to the spot rate through dealer hedging.

JEL Codes: F31, G12, G15, G23, D83

Keywords: Exchange rate determination, FX forwards, Covered interest parity, OTC markets

*Aldasoro: Bank for International Settlements inaki.aldasoro@bis.org. Calani: Central Bank of Chile, mcalani@bcentral.cl. Romero: Central Bank of Chile, dromeroc@bcentral.cl. The views expressed are those of the authors and do not necessarily reflect those of the Bank for International Settlements, the Central Bank of Chile (CBC) or its Board members. This study was developed within the scope of the research agenda conducted by the CBC in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities, by virtue of collaboration agreements signed with these institutions.

1 Introduction

Foreign exchange (FX) forward markets are central to currency risk management and price formation, especially in emerging economies where a large share of forward trading occurs over-the-counter (OTC). In these markets, clients face search frictions, quotes are negotiated bilaterally, and dealers' intermediation capacity is constrained by balance-sheet, funding, and collateral considerations. When dealer balance-sheet capacity becomes scarce (due to congestion, margin requirements, or heightened financial stress), forward prices can deviate from parity benchmarks such as covered interest parity (CIP). These deviations can then feed back into the spot rate through dealers' inventory hedging in spot and bond markets. Therefore, OTC frictions affect the pricing of currency risk, the ability of institutions to hedge foreign-currency exposures, and the transmission of global financial conditions to exchange rates and domestic financial variables, especially in episodes of stress when balance-sheet and margin constraints bind more tightly.

This paper proposes a tractable framework to study how OTC frictions shape forward pricing, the forward basis, and spot exchange rate determination in equilibrium. The model features two blocks. The competitive block consists of spot FX and domestic and foreign risk-free bonds that trade competitively. The OTC block features bilateral meetings between clients and dealers and bargaining over forward quotes in an alternating-offers protocol with endogenous breakdown risk. On the demand side, we allow for a general class of mean–variance clients that nests natural hedgers with exogenous foreign-currency exposures and global investors with purely speculative demand. In both cases, optimal forward demand is linear in the negotiated quote and decomposes cleanly into hedging (intercept) and speculative (slope) components. On the supply side, dealers intermediate client trades, hedge their net forward inventory using the foreign bond, and face convex intermediation costs whose curvature increases with market tightness (clients per dealer) and an aggregate stress state.

Our results show that equilibrium forward quotes, the basis, and the spot exchange rate are jointly determined by (i) the allocation of bargaining power in OTC negotiations, (ii) endogenous congestion (market tightness) that shifts intermediation costs and outside options, and (iii) dealers' hedging feedback into competitive spot and bond markets. As a result, stress-driven balance-sheet scarcity can generate sizable departures from parity, which can be amplified (or attenuated) through dealers' inventory hedging.

The key mechanism is intuitive. When the OTC market becomes more congested (higher tightness), clients face worse meeting prospects and weaker bargaining positions, while dealers can more readily reallocate scarce capacity across counterparties. This reduces participation, disciplines traded volumes, and pushes negotiated forward quotes toward clients' reservation rates (i.e., away from dealers' break-even rates), as dealers extract a larger share of the match surplus. When financial stress rises, effective balance-sheet curvature increases, amplifying the marginal cost of intermediation and widening the basis. Because dealers hedge their aggregate forward inventory in the foreign-bond market, these forces also shift their bond demand and, in turn, the equilibrium spot rate. The result is a unified theory of OTC frictions, the forward basis, and the spot exchange rate, with sharp comparative statics for prices and quantities.

In ongoing work, we (i) discuss how the framework can be taken to the data to better characterize the magnitude of the frictions and the cross-sectional heterogeneity in client demand, (ii) study episodes of CIP stress, and (iii) analyze policy-relevant changes in intermediation capacity and policy interventions.

The paper speaks to three strands of literature. First, it connects to work on CIP deviations and limits to arbitrage, which emphasizes the role of balance-sheet and funding constraints in intermediary pricing (Borio et al., 2016; Du et al., 2018; Avdjiev et al., 2019; Rime et al., 2022). Second, it relates to FX microstructure research highlighting dealer inventory, risk management, and intermediation capacity as determinants of quotes and liquidity (Evans and Lyons, 2002; Gabaix and Maggiori, 2015; Huang et al., 2025). Third, it builds on OTC search-and-bargaining frameworks in the Rubinstein (1982) tradition, where market tightness and outside options shape within-match surplus splitting and thus transaction prices (Duffie et al., 2005; Lagos and Rocheteau, 2009; Lagos et al., 2011; Atkenson et al., 2015). Our framework combines these ingredients to deliver (i) an endogenous forward rate pinned down by bargaining under tightness, (ii) an endogenous basis arising from state- and tightness-dependent intermediation costs, and (iii) equilibrium feedback to the spot rate through dealers' hedging demand in the competitive bond and spot markets.

The remainder of the paper is organized as follows. Section 2 presents the model and characterizes individual demand, dealer supply, search and bargaining, within-match outcomes, and the determination of market tightness in equilibrium. Section 3 (to be completed) characterizes the equilibrium, discusses quantitative implications and empirical mapping of tightness and stress, and outlines how the model can be disciplined using FX forward data and position/hedging information. Section 4 concludes.

2 A Model of Spot and Forward Exchange Rate Determination

In this section we develop a model of spot and forward exchange rate determination in a semi-decentralized exchange rate (FX) market. The key idea is that the spot FX rate is determined in a competitive Walrasian block, while over-the-counter (OTC) forward contracts between clients and a dealer are subject to search frictions and bilateral bargaining. On the client side we consider a general type of agents that can characterize individuals who have exogenous foreign-currency exposures they wish to insure (natural hedgers), or agents who take forward positions to exploit perceived risk–return trade-offs (global investors). On the dealer side, a competitive fringe of intermediaries takes the opposite side of client orders, hedges in the underlying bond and spot FX markets, and incurs CIP-deviation costs when carrying hedged positions. The interaction between client demand, dealer hedging, and search frictions in the OTC segment generates an endogenous forward rate, an equilibrium basis, and feedback to the spot FX rate.

Time is discrete with two dates, $t = 1, 2$. At $t = 1$ agents trade domestic and foreign risk-free bonds with deterministic gross returns R^d and R^f, respectively, where the spot FX is S_1 (domestic currency per unit of foreign currency). In parallel, clients and dealers enter an OTC forward market where they negotiate a forward rate F for delivery of one unit of foreign currency at $t = 2$, when the spot rate S_2 is realized and all payoffs are settled. The only source of risk is the spot FX in the second period.

Within the OTC block, negotiations at $t = 1$ can take multiple rounds indexed by $n = 0, 1, 2, \ldots$ before an agreement is reached or bargaining breaks down. We allow for heterogeneous discount factors per round (within $t = 1$) across clients and dealers, which are denoted by δ_C and δ_D, respectively.

We first describe the preferences and optimal demands of individual agents in a generic mean–variance framework, then introduce the search and matching environment, value functions, and bargaining problem that jointly determine F as a function of market tightness and outside options. Finally, we close the model by imposing Walrasian market clearing in the foreign-bond and spot FX markets, which pins down S_1 and links the forward basis to the composition of OTC client demand and dealer balance-sheet constraints.

2.1 Demand of Forward Contracts

We consider a generic client indexed by $i \in \mathcal{I}$. A type-i is summarized by a risk-aversion parameter $\gamma_i > 0$ and an exogenous foreign-currency exposure $\widetilde{X}_i$ (which can be positive, negative, or zero). All payoffs and utility are measured in domestic currency. Clients have mean–variance preferences over terminal wealth $W_{2,i}$, $U_i(W_{2,i}) = \mathrm{E}[W_{2,i}] - \frac{\gamma_i}{2}\mathrm{var}(W_{2,i})$.

At $t = 1$ the client trades an OTC FX forward at rate F (domestic currency per unit of foreign currency) with a *signed* notional X (units of foreign currency). We adopt the convention that $X > 0$ means the client sells foreign currency forward (delivers X units of foreign currency at $t = 2$ and receives FX units of domestic currency), while $X < 0$ means the client buys foreign currency forward.

The client also has an exogenous exposure $\widetilde{X}_i$ in foreign currency maturing at $t = 2$. We allow this exposure to earn a gross return R^f in foreign currency between $t = 1$ and $t = 2$. Hence, the domestic-currency value of the exposure at $t = 2$ is $\widetilde{X}_i R^f S_2$. The forward leg adds the domestic payoff $X(F - S_2)$ at $t = 2$. Up to deterministic terms unrelated to X, terminal wealth can be written as $W_{2,i}(X) = \widetilde{X}_i R^f S_2 + X(F - S_2)$.

Because the only source of risk is S_2, the client chooses X to trade off mean and variance of the net S_2 exposure. Letting $\mu \equiv \mathrm{E}[S_2]$ and $\sigma^2 \equiv \mathrm{var}(S_2)$, the client solves

$$\max_{X\in\mathbb{R}} \mathrm{E}\Big(\widetilde{X}_i R^f S_2 + X(F - S_2)\Big) - \frac{\gamma_i}{2}\mathrm{var}\Big(\widetilde{X}_i R^f S_2 + X(F - S_2)\Big).$$

The first-order condition yields the optimal signed notional demand:

$$X_i^*(F) = \widetilde{X}_i R^f + \frac{F - \mathrm{E}[S_2]}{\gamma_i\,\mathrm{var}(S_2)}. \tag{1}$$

This demand decomposes into (i) a *hedging/intercept* component $\widetilde{X}_i R^f$ and (ii) a *speculative/slope* component proportional to the forward mispricing $F - \mathrm{E}[S_2]$. Importantly, $X_i^*(F)$ can be positive or negative depending on $(\widetilde{X}_i, \gamma_i, F)$.

Natural hedgers and global investors as special cases. Under the generic problem above, "natural hedgers" correspond to types with $\widetilde{X}_i \neq 0$, while "global investors" correspond to types with $\widetilde{X}_i = 0$. In particular, when $\widetilde{X}_i \neq 0$, we can express choices in terms of a hedging ratio $h_i \equiv X/(\widetilde{X}_i R^f)$

$$h_i^*(F) \equiv \frac{X_i^*(F)}{\widetilde{X}_i R^f} = 1 - \frac{\mathrm{E}[S_2] - F}{\gamma_i\,\mathrm{var}(S_2)\,\widetilde{X}_i R^f},$$

which delivers the same expressions presented by Chen and Zhou (2025) and Liao and Zhang (2025). Importantly, this optimal hedging ratio imposes no constraints regarding the behavior of the client.

On the contrary, when $\widetilde{X}_i = 0$, Eq. (1) reads as

$$X_i^*(F) = \frac{F - \mathrm{E}[S_2]}{\gamma_i\,\mathrm{var}(S_2)},$$

which is the speculative demand of an agent who has no pre-existing FX exposure but can take exposure through forwards. These agents are similar to the speculators in Chen and Zhou (2025) or the global investors in De Leo et al. (2025).

Our previous formulation provides a general framework for determining the demand for forward contracts, regardless of the motives behind them or their nature (selling or buying). We now characterize the supply side of these contracts.

2.2 Supply of Forward Contracts

We assume there are domestic intermediaries with access to domestic and foreign bonds, as well as to domestic spot and forward FX markets. These dealers are willing to provide forward contracts to clients as long as it is profitable to do so. However, and to capture the nature of intermediation in emerging economies (Alfaro et al., 2023; De Leo et al., 2025), we assume they face two constraints. First, these intermediaries cannot have "open FX positions", implying they must fully hedge their FX risk in every period. Second, they face position limits and risk-bearing capacity constraints.

To capture these ideas in a simple way, we make two assumptions. First, we assume that dealers engage in a fully-hedged zero-capital FX arbitrage strategy of signed scale $K \in \mathbb{R}$. For $K > 0$, at $t = 1$ a dealer borrows K in foreign currency at rate R^f, converts it to domestic currency at the spot rate S_1 and invests it in the domestic bond for a return R^d; she then hedges the foreign repayment KR^f by buying that amount forward at rate F. For $K < 0$, the same expression represents the reverse strategy (borrowing domestic, lending foreign, and selling forward). In either case, the signed forward notional is $X = KR^f$ and the total profit from this strategy is $\Pi_D(F) = K\big(S_1 R^d - FR^f\big)$. Second, we assume the dealer faces a quadratic balance-sheet cost for carrying the hedged position of scale K, $\text{Cost}(K) = \frac{\kappa(\theta,\omega)}{2}K^2$, with $\kappa(\theta, \omega) > 0$.[1]

Marginal costs of intermediation. The parameter $\kappa(\theta, \omega)$ governs the marginal convexity of balance-sheet costs, so a larger κ means that scaling the dealer's intermediation capacity becomes more expensive at the margin. Economically, higher κ captures tighter internal risk limits, more expensive funding and collateral/margin requirements, regulatory capital constraints, and higher opportunity costs of tying up scarce dealer resources.

Importantly, we allow this marginal cost to depend on two aggregate conditions. First, on market tightness θ (clients per dealer), which we discuss in greater detail below. Second, intermediation costs depend on the degree of stress in the market, summarized by ω. We assume that marginal costs are increasing in both arguments, so $\frac{\partial\kappa(\theta,\omega)}{\partial\theta} > 0$ and $\frac{\partial\kappa(\theta,\omega)}{\partial\omega} > 0$ hold. This assumption captures the idea that it is more costly to provide intermediation when dealers face a larger mass of clients or when the market is under stress.[2]

Assumption 1 (Congestion and stress increase balance-sheet convexity)**.** *The curvature $\kappa(\theta, \omega)$ is continuously differentiable and satisfies $\frac{\partial\kappa(\theta,\omega)}{\partial\theta} > 0$ and $\frac{\partial\kappa(\theta,\omega)}{\partial\omega} > 0$.*

[1] Our assumption that dealers fully hedge their FX exposure is standard and ensures that dealers' only source of profit is the bid-ask spread net of balance-sheet costs. Relaxing this assumption to allow dealers to take selective speculative positions would introduce an additional layer of complexity as dealers would then trade off expected profits from unhedged positions against the costs of bearing FX risk. We maintain the full-hedging assumption for tractability, noting that it captures the core friction–balance-sheet costs–while keeping the analysis sharp.

[2] We think about ω as capturing several empirically and theoretically plausible notions of stress. For example, a latent funding/capital tightness factor (a shadow value of balance sheet costs), an observable proxy for funding spreads, or global risk indices which are perceived as exogenous from the perspective of the small open economy, or a discrete regime shift from normal to stress periods.

The critical observation is to note that, even if each dealer has a fixed "physical" risk-management technology, the *shadow cost* of deploying balance sheet depends on market-wide conditions. In particular on (i) the intensity of client demand per dealer (summarized by θ) and (ii) the funding/capital environment (summarized by ω). Thus, $\kappa(\theta, \omega)$ is best viewed as a reduced-form mapping from aggregate states into the marginal convexity of balance-sheet usage. In Appendix A.1 we provide further motivations for the dependence of κ on aggregate conditions.

Profits from intermediation. The net profit per K units of capacity from this position, in domestic currency, is

$$\Pi(F, K, \theta, \omega) = K\big(S_1 R^d - R^f F\big) - \frac{\kappa(\theta, \omega)}{2} K^2.$$

If agreeing to intermediate notional X, the demand for forward contracts pins down the hedged scale via $KR^f = X$ (so $K = X/R^f$). Then, net profit becomes

$$\begin{aligned} \Pi(F, X, \theta, \omega) &= \frac{X}{R^f}\left(S_1 R^d - R^f F\right) - \frac{\kappa(\theta, \omega)}{2}\left(\frac{X}{R^f}\right)^2 \\ &= X\left(\frac{S_1 R^d}{R^f} - F\right) - \frac{\kappa(\theta, \omega)}{2}\left(\frac{X}{R^f}\right)^2. \end{aligned}$$

Equivalently, $\Pi(F, X, \theta, \omega)$ can be interpreted as the dealer's profit from taking the opposite side of a client forward position of signed notional X and hedging the resulting FX exposure competitively in spot and bond markets at prices (S_1, R^d, R^f).

In a Walrasian setup, the demand and supply for forward contracts would be enough to determine a *unique* forward rate as function of deep parameters, the spot rate and interest rates. In what follows, we depart from the literature to characterize the OTC nature of the forward market.

2.3 Search, Bargaining and Forward Rate Determination

We now describe the search and bargaining frictions in the OTC forward market. Fig. 1 summarizes the main ingredients and stages of the model. Fix an aggregate state (S_1, ω). Clients with FX exposures and dealers supplying balance-sheet capacity meet bilaterally at random, bargain over the forward rate, and possibly break off negotiations. The key objects are (i) the matching probabilities implied by an aggregate matching function, (ii) the within-match breakdown probabilities $\lambda_C(\theta)$ and $\lambda_D(\theta)$ as functions of market tightness, (iii) the bargaining weight $\alpha^{(p)}(\theta)$ induced by the bargaining game, and (iv) the type-i unmatched value $J_i^U(\theta, \omega)$ that governs participation.

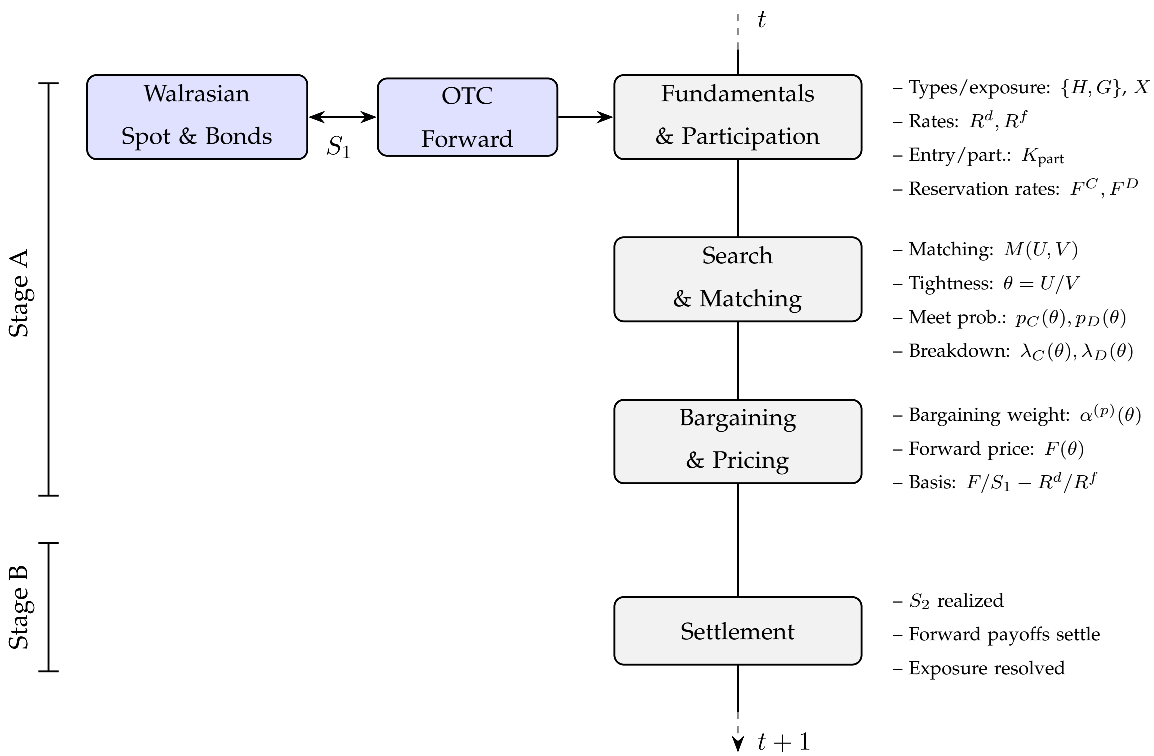


Figure 1. Timing and Structure of the FX Market

2.3.1 Matching Technology and Tightness

Let U denote the mass of clients in the OTC forward market and V the mass of dealers. Meetings between clients and dealers are governed by an aggregate matching function $M(U, V)$ satisfying standard assumptions.[3]

We define the market tightness as $\theta \equiv \frac{U}{V}$ (mass of clients per dealer), so a higher θ means a more congested, client-heavy market, while a lower θ means a dealer-rich market. Due to constant returns to scale, $M(U, V) = Vm(\theta)$ holds. We define the per-round matching probabilities for a generic client and dealer as

$$p_C(\theta) \equiv \frac{M(U,V)}{U} = \frac{m(\theta)}{\theta}, \tag{2}$$

$$p_D(\theta) \equiv \frac{M(U,V)}{V} = m(\theta). \tag{3}$$

Because the matching function is increasing and homogeneous of degree one, $p_C'(\theta) < 0$ and $p_D'(\theta) > 0$ hold. Intuitively, when there are more clients per dealer (higher θ), clients find dealers less easily (the opposite for dealers). The next assumption summarizes the characteristics of the meeting probabilities.

[3] M is continuously differentiable, strictly increasing in each argument, and homogeneous of degree one (constant returns to scale). We also impose $M(0, V) = M(U, 0) = 0$ and $M(U, V) \leq \min\{U, V\}$.

Assumption 2 (Monotone meeting probability)**.** *The client and dealer meeting probabilities* $p_C(\theta)$ *and* $p_D(\theta)$ *are continuously differentiable and strictly decreasing/increasing, respectively,* $p_C'(\theta) < 0$ *and* $p_D'(\theta) > 0$.

2.3.2 Reservation Rates and Surplus

For a given *signed* foreign-currency notional X, we define the reservation forward rates of the type-i client and the dealer as F_i^C and $F^D(X;\theta,\omega)$, respectively. These are the forward rates at which each party is indifferent between trading and not trading. For any forward rate F and notional X, the incremental certainty-equivalent surpluses from signing a forward contract are

$$\text{Surplus}_{C,i}(F) = X(F - F_i^C),$$
$$\text{Surplus}_D(F) = X\big(F^D(X;\theta,\omega) - F\big).$$

Note that these definitions work uniformly for $X > 0$ and $X < 0$. When $X > 0$ (client sells foreign forward), the minimum forward rate at which the client is willing to trade is F_i^C and the client surplus increases in F. On the other hand, the dealer surplus decreases in F, with a maximum acceptable forward rate of $F^D(X;\theta,\omega)$, so trade requires $F^D(X;\theta,\omega) > F_i^C$. When $X < 0$ (client buys foreign forward), the monotonicities reverse, and trade requires $F^D(X;\theta,\omega) < F_i^C$. In both cases, the gains-from-trade condition requires total surplus to be positive,

$$S_i(X;\theta,\omega) \equiv \text{Surplus}_{C,i}(F) + \text{Surplus}_D(F) = X\big(F^D(X;\theta,\omega) - F_i^C\big) > 0,$$

which is independent of F as long as trade occurs.

Considering the agents described in Section 2.2, the reservation rates are as follows. For a type-i client, the reservation rate is pinned down by the condition $X_i^*(F_i^C) = 0$. Using Eq. (1), we obtain

$$F_i^C = \mathrm{E}[S_2] - \gamma_i \mathrm{var}(S_2)\widetilde{X}_i R^f,$$

which is independent of the signed notional X. If $\widetilde{X}_i > 0$ (e.g., an exporter), F_i^C is the minimum forward rate at which the client is willing to *sell* foreign currency forward. If $\widetilde{X}_i < 0$ (e.g., an importer), F_i^C is the maximum forward rate at which the client is willing to *buy* foreign currency forward. When $\widetilde{X}_i = 0$ (a "global investor" in the sense of Section 2.1), we have $F_i^C = \mathrm{E}[S_2]$, so the reservation forward rate equals the expected future spot rate and does not depend on γ_i.

For the dealer, we have that the break-even forward rate that makes her indifferent between trading or not is defined by the condition $\Pi(F^D(X;\theta,\omega), X, \theta, \omega) = 0$, implying $F^D(X;\theta,\omega) = F^{\text{CIP}} - \frac{\kappa(\theta,\omega)}{2}\frac{X}{(R^f)^2}$, with $F^{\text{CIP}} \equiv \frac{S_1 R^d}{R^f}$ denoting the covered interest parity benchmark (CIP) in domestic currency per unit of foreign currency. Note that, because the dealer takes the client's position as given, this delivers a tightness- and state-dependent reservation forward rate and thus a tightness- and state-dependent match surplus.

Fixing an aggregate state (S_1, ω) and conjecturing a degree of market tightness θ, conditional on meeting a dealer, a type-i client and the dealer bargain. Because (i) dealer intermediation costs depend on the notional size through balance-sheet curvature and (ii) the client's desired notional depends on the negotiated quote, the within-match outcome is a joint determination of *price and quantity*. We formally characterize this bargaining process in what follows.

2.3.3 Within-Match Bargaining

Following Rubinstein (1982) and Binmore et al. (1986), we model the negotiation for a forward rate as an infinite-horizon alternating-offers game (strategic bargaining game).

This game can be characterized as follows. There are n bargaining rounds, with $n = 0, 1, 2, \ldots$, all taking place within $t = 1$, and in each round one party proposes a forward rate F_n. We consider two protocols: client-first ($p = CF$) and dealer-first ($p = DF$). In the client-first protocol, the client proposes at $n = 0$ and the dealer responds; roles then alternate for $n = 1, 2, \ldots$. In the dealer-first protocol, the dealer proposes at $n = 0$ and roles again alternate thereafter. Both parties discount future payoffs geometrically between bargaining rounds, with $\delta_C \in (0, 1)$ for the client and $\delta_D \in (0, 1)$ for the dealer. We interpret δ_C and δ_D as per-round (within–$t = 1$) discount factors capturing opportunity costs of delay and the risk of losing the match. The relevant time unit for being unmatched is therefore the bargaining round.

After an offer is made, the recipient can accept or reject. If the offer is accepted, the contract is signed at the proposed rate and the game ends. The client and the dealer receive surpluses $\text{Surplus}_{C,i}(F)$ and $\text{Surplus}_D(F)$, which sum to the total match surplus. However, if the offer is rejected, the match may break down with a certain probability to be described below. If the relationship ends, both parties return to the unmatched state. Since $\text{Surplus}_{C,i}(F)$ and $\text{Surplus}_D(F)$ are defined as incremental surpluses relative to not trading (or equivalently, relative to returning unmatched), we normalize the payoff upon breakdown to zero in this surplus-splitting problem. If the relationship survives, the game proceeds to the next round with roles reversed.

We assume that the breakdown probability depends on whose offer was rejected, which we interpret as whose outside option is activated by being turned down. If the rejected offer was made by the client, the match breaks with probability $\lambda_C(\theta)$. This captures the notion of a client-driven shop-around: after her offer is rejected, the client may leave the table to search elsewhere. On the other hand, if the rejected offer was made by the dealer, the match breaks with probability $\lambda_D(\theta)$. This probability captures a dealer-driven congestion, implying that after her quote is rejected, the dealer may abandon the negotiation to reallocate scarce balance-sheet capacity. In what follows, we focus on the empirically plausible case in which $\lambda'_C(\theta) < 0$ and $\lambda'_D(\theta) > 0$ hold. That is, in a more congested market (higher θ) there are many clients per dealer, so it is less likely that a rejected client finds alternative offers outside. At the same time, if an offer made by a dealer is rejected, it is more likely to find alternative clients outside. In Appendix A.2 we provide microfoundations for these breakdown probabilities and their link to the matching probabilities described above.

Assumption 3 (Asymmetric-breakdown probabilities). *The breakdown functions are continuously differentiable and satisfy $\lambda'_C(\theta) < 0$ and $\lambda'_D(\theta) > 0$ for all θ.*

We are now ready to solve for the equilibrium forward rate derived from this game. We solve it under different bargaining protocols where clients and dealers can make the first offer. We summarize our results in the following proposition.

Proposition 1. *Consider the search-and-bargaining OTC FX model in which, conditional on a match, a type-i client and a dealer bargain over a forward rate. Let $\delta_C, \delta_D \in (0,1)$ be their per-round discount factors and let $\lambda_C(\theta), \lambda_D(\theta) \in (0,1)$ denote the breakdown probabilities in a bargaining round, where θ is the market tightness (clients per dealer). Let F_i^C and $F^D(X;\theta,\omega)$ denote their reservation rates. The equilibrium forward rate satisfies*

$$F_i(\theta,\omega) = (1-\alpha^{(p)}(\theta))F_i^C + \alpha^{(p)}(\theta)F^D\big(X_i(\theta,\omega);\theta,\omega\big), \tag{4}$$

where $\alpha^{(p)}(\theta)$ is the client's equilibrium share of the bilateral surplus under protocol $p \in \{CF, DF\}$.

Denoting the continuation probabilities as $q_C(\theta) = 1-\lambda_C(\theta)$ and $q_D(\theta) = 1-\lambda_D(\theta)$, when client offers first ($p = CF$), her equilibrium share of the total match surplus is

$$\alpha^{CF}(\theta) = \frac{1-q_C(\theta)\delta_D}{1-q_C(\theta)q_D(\theta)\delta_C\delta_D} \in (0,1), \tag{5}$$

while when the dealer offers first ($p = DF$), the client's share of the total match surplus is

$$\alpha^{DF}(\theta) = q_D(\theta)\delta_C\alpha^{CF}(\theta) = q_D(\theta)\delta_C\frac{1-q_C(\theta)\delta_D}{1-q_C(\theta)q_D(\theta)\delta_C\delta_D} \in (0,1). \tag{6}$$

Proof. See Appendix A.3.1. □

Proposition 1 states the familiar result derived from this class of games applied to the OTC FX forward market. The equilibrium forward rate within the match corresponds to a weighted average of the reservation rates of each party. Conditional on a bargaining protocol, the weight $\alpha^{(p)}(\theta)$ depends on (i) the discount factors of each party and (ii) the market tightness through the breakdown probabilities.

In the standard strategic bargaining game, those breakdown probabilities are nil ($\lambda_C(\theta), \lambda_D(\theta) \to 0$), the sharing rule only depends on the relative discount factors, and the degree of market tightness is irrelevant to determine the outcome of the game. In such framework, there are two well-known properties of the equilibrium: (i) the more patient agent (the one with higher δ) tends to have a higher share of the surplus and (ii) there is a first-mover advantage.

In our case, the introduction of the breakdown probabilities and their dependence on the market tightness generates sharp predictions regarding the client's equilibrium share. The next proposition characterizes the effect of market tightness on bargaining outcomes.

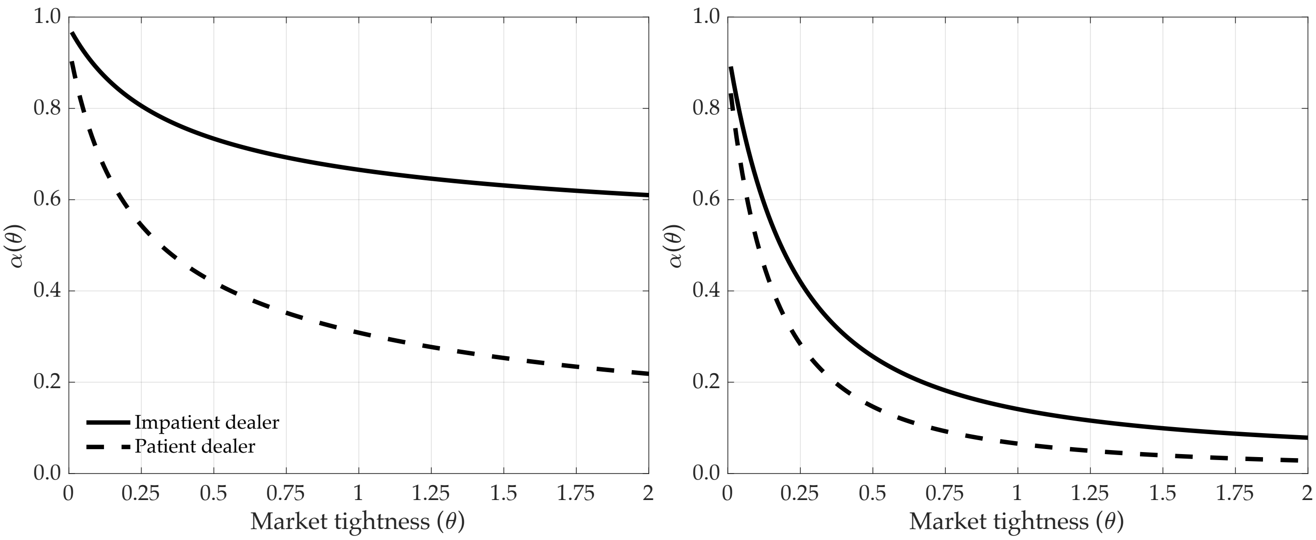
Proposition 2. *Under Assumption 3, the client's bargaining weight $\alpha^{(p)}(\theta)$ is strictly decreasing in tightness, for any bargaining protocol.*

Proof. See Appendix A.3.2. □

The intuition behind Proposition 2 is as follows. By Assumption 3, $\lambda'_C(\theta) < 0$ and $\lambda'_D(\theta) > 0$ hold, so an increase in tightness θ simultaneously (i) makes the client less willing to walk away after her offer is rejected (stabilizing the relationship in a way that favors the dealer) and (ii) makes the dealer more willing to walk away after his quote is rejected (weakening the client's ability to reject and wait). Both forces reduce the client's equilibrium share, regardless of who moves first. Dealer-first adds an additional multiplicative dampening factor $q_D(\theta)$. When dealers are more prone to abandon negotiations (high λ_D), the client captures an even smaller share.

The critical implication of the proposition is that the client's share unambiguously decreases in market tightness and, even though the first-mover advantage still exists, it plays a minor role in the determination of the forward rate. Fig. 2 provides examples of the client's share rule under different bargaining protocols and different levels of dealer impatience. Consistent with the previous intuition, regardless of the bargaining protocol, the client's share monotonically decreases in market tightness, even though the level is higher when the client can offer first. Also notice that the more patient the dealer is, the lower the client's share.

Proposition 1 characterizes the equilibrium forward quote within a match $F_i(\theta, \omega)$ as a convex combination of the client and dealer reservation rates. At the same time, the client's requested notional must be optimal at the negotiated quote, $X_i(\theta, \omega) = X_i^*\big(F_i(\theta, \omega)\big)$, with $X_i^*(\cdot)$ given by Eq. (1). Those two conditions jointly pin down, for each type-i, a within-match equilibrium pair $(F_i(\theta, \omega), X_i(\theta, \omega))$. The sign of $X_i(\theta, \omega)$ is therefore an equilibrium outcome that depends on both type characteristics (e.g., $\widetilde{X}_i, \gamma_i$) and market conditions (through $\alpha^{(p)}(\theta)$ and $\kappa(\theta, \omega)$). The next result summarizes the equilibrium within a match.



(a) Client Offers First

(b) Dealer Offers First

Figure 2. Client's Bargaining Weight as a Function of Market Tightness

Notes: This figure plots the client's equilibrium bargaining weight in the OTC FX model under different bargaining protocols. The figures are produced assuming asymmetric breakdown probabilities with Poisson arrival rates, $\lambda_i(\theta) = 1 - \exp(-\nu_i p_i(\theta))$, where $p_i(\theta)$ is the matching probability coming from the function $M(U, V) = \frac{UV}{U+V}$ (den Haan et al., 2000). The curvature exponents are $\nu_C = 0.5$ and $\nu_D = 3$. The client's discount factor is $\delta_C = 0.95$. Impatient dealer corresponds to $\delta_D = 0.50$. Patient dealer corresponds to $\delta_D = 0.95$.

The linear-quadratic structure of our model yields a closed-form solution for the within-match equilibrium. This tractability allows us to characterize explicitly how tightness and stress affect prices and quantities, as summarized in the following corollary.

Corollary 1 (Within-match equilibrium)**.** *Because X_i^* is linear in F_i, and F^D is linear in X_i, the within-match equilibrium admits the following closed form*

$$X_i(\theta, \omega) = \frac{\alpha^{(p)}(\theta)}{\gamma_i \operatorname{var}(S_2) + \frac{\alpha^{(p)}(\theta)\kappa(\theta,\omega)}{2(R^f)^2}} \big(F^{\text{CIP}} - F_i^C\big), \tag{7}$$

$$F_i(\theta, \omega) = F_i^C + \alpha^{(p)}(\theta) \frac{\gamma_i \operatorname{var}(S_2)}{\gamma_i \operatorname{var}(S_2) + \frac{\alpha^{(p)}(\theta)\kappa(\theta,\omega)}{2(R^f)^2}} \big(F^{\text{CIP}} - F_i^C\big). \tag{8}$$

Eqs. (7) and (8) show that type-i trading volume and pricing move with tightness only through $\alpha^{(p)}(\theta)$ and $\kappa(\theta, \omega)$. By Proposition 2, higher tightness lowers $\alpha^{(p)}(\theta)$, while higher balance-sheet curvature $\kappa(\theta, \omega)$ reduces volume and shifts the negotiated quote toward F_i^C.

2.3.4 Value Function and Equilibrium Market Tightness

We now embed this bilateral bargaining block into a stationary search environment defined over bargaining rounds within $t = 1$. Let $J_i^U(\theta, \omega)$ denote the expected discounted surplus of a representative type-i client that is currently unmatched in the OTC forward market. We consider a stationary equilibrium with constant tightness θ.

In each bargaining round, an unmatched client meets a dealer with probability $p_C(\theta)$ and remains unmatched with probability $1 - p_C(\theta)$. Conditional on a match, the client trades at the within-match equilibrium and obtains incremental surplus $\Sigma_{C,i}^{(p)}(\theta, \omega) = \alpha^{(p)}(\theta) S_i(\theta, \omega)$, where $S_i(\theta, \omega) \equiv S_i\big(X_i(\theta, \omega); \theta, \omega\big)$ is the total match surplus evaluated at the equilibrium notional. Under stationarity, after a trade the client returns to the unmatched state in the next bargaining round (equivalently, the matched client is replaced by an identical unmatched client). The Bellman equation for the unmatched client value is therefore

$$J_i^U(\theta, \omega) = \delta_C \Big(p_C(\theta) \big[J_i^U(\theta, \omega) + \Sigma_{C,i}^{(p)}(\theta, \omega) \big] + (1 - p_C(\theta)) J_i^U(\theta, \omega) \Big).$$

The term in brackets reflects that, upon matching, the client receives the surplus $\Sigma_{C,i}^{(p)}(\theta, \omega)$ and then starts the next round in the same unmatched state due to stationarity.

Rearranging the Bellman equation yields

$$J_i^U(\theta, \omega) = \frac{\delta_C}{1 - \delta_C} p_C(\theta)\, \Sigma_{C,i}^{(p)}(\theta, \omega) = \frac{\delta_C}{1 - \delta_C} p_C(\theta)\, \alpha^{(p)}(\theta)\, S_i(\theta, \omega).$$

Note that the sign of X is an *intensive-margin* outcome that determines the direction and size of trade conditional on meeting a dealer and therefore affects within-match surplus and J_i^U. By contrast, market tightness θ is an *extensive-margin* outcome that determines how many clients choose to participate in the OTC market relative to the dealer mass; it affects bargaining and dealer costs and feeds back into (F_i, X_i) and the resulting surplus. In our environment, all participating clients, regardless of whether their optimal X is positive or negative, search in the *same* OTC market and meet from the *same* pool of dealers. Therefore, there is a single matching congestion variable θ, which we now characterize.

Entry, participation, and the determination of θ. Let k denote an idiosyncratic participation cost drawn from the cumulative distribution function (CDF) G, where $G(x) \equiv \Pr(k \leq x)$. This CDF is common across client types and determines the probability of participating in the forward market whenever the gains from doing so exceed the cost, $J_i^U(\theta, \omega) \geq k$. Formally, $\Pr(\text{participate} \mid \theta, \omega, i) = \Pr\big(J_i^U(\theta, \omega) \geq k\big) = G\big(J_i^U(\theta, \omega)\big)$.

We now characterize the existence and uniqueness of market tightness, conditional on the aggregate state (S_1, ω). We make the following assumption.

Assumption 4 (Regularity and monotonicity of entry values)**.** *Fix ω and treat S_1 as given (equivalently, take F^{CIP} as fixed). For each $i \in \mathcal{I}$, the unmatched value $J_i^U(\theta, \omega)$ is well-defined for $\theta \geq 0$, continuous in θ, and weakly decreasing in θ. The participation-cost CDF $G(\cdot)$ is weakly increasing and continuous.*

In the mean–variance environment of Section 2.1, the monotonicity in Assumption 4 follows from primitive monotonicities of the matching and bargaining blocks (and from $\frac{\partial \kappa(\theta,\omega)}{\partial \theta} > 0$); see Appendix A.3.3. Hence, Assumption 4 is not restrictive.

We consider the set of clients to be discrete ($\mathcal{I} = \{1, \ldots, I\}$) and let $\overline{U}_i$ denote the potential mass of type-i clients. Given (θ, ω), the endogenous active mass of type-i is $U_i(\theta, \omega) = \overline{U}_i\, G\big(J_i^U(\theta, \omega)\big)$, while the total mass of active clients is $U(\theta, \omega) = \sum_{i=1}^{I} U_i(\theta, \omega)$. Then, equilibrium market tightness solves the fixed point

$$\theta = T(\theta; \omega) \equiv \frac{1}{V} \sum_{i=1}^{I} \overline{U}_i\, G\big(J_i^U(\theta, \omega)\big). \tag{9}$$

The next proposition summarizes the existence and uniqueness of equilibrium tightness in the forward OTC market, holding the aggregate state fixed.

Proposition 3 (Existence and uniqueness of equilibrium tightness)**.** *Suppose Assumption 4 holds and $\sum_{i=1}^{I} \overline{U}_i < \infty$. Then, for any fixed aggregate state (S_1, ω) (equivalently, for any fixed (F^{CIP}, ω)), the fixed-point Eq. (9) admits a unique solution $\theta^* = \theta(S_1, \omega) \in [0, \bar{\theta}]$, where $\bar{\theta} \equiv \frac{1}{V} \sum_{i=1}^{I} \overline{U}_i$.*

Proof. See Appendix A.3.3. □

The mapping $T(\theta; \omega)$ captures the endogenous mass of participating clients. At a given tightness, each type-i client enters with probability $G(J_i^U(\theta, \omega))$, which is smaller when tightness is higher because congestion reduces meeting chances and bargaining outcomes (and increases dealer costs). Hence $T(\theta; \omega)$ is decreasing, while the $45°$ line θ is increasing, delivering a unique intersection. Relative to a baseline representative-client free-entry condition, heterogeneous costs replace a knife-edge indifference condition with an endogenous participation schedule; the OTC market may still be inactive, but this corresponds to a fixed point at $\theta^* = 0$ rather than to non-existence.

With these ideas in hand, it is useful to distinguish three related notions. First, the OTC market can be *inactive* on the extensive margin if the fixed point delivers $\theta^* = 0$, which implies $U(\theta^*, \omega) = 0$ and thus zero meetings and trades. Second, even when the market is active, some types may have *no trade* on the intensive margin. For a given i, the within-match equilibrium may satisfy $X_i(\theta^*, \omega) = 0$ (equivalently, $S_i(\theta^*, \omega) = 0$), so that type is indifferent and does not trade conditional on meeting.[4] Third, aggregate

[4] In our linear–quadratic environment, $X_i(\theta, \omega) = 0$ if and only if $F^{\text{CIP}} = F_i^C$, so the dealer benchmark quote coincides with the client's reservation rate.

net dealer inventory may be zero even when the OTC market is active: because notionals are signed, offsetting client positions can imply $\mathcal{X}^{\text{OTC}}(\theta^*, \omega) = 0$ despite positive gross volume. This matters for the general-equilibrium feedback to the spot rate, since dealers hedge only their *net* forward inventory in the Walrasian bond market. We now turn to aggregation.

2.4 Aggregate OTC Demand and Dealer Inventory

To highlight the general-equilibrium link, we now make explicit the dependence on the spot rate S_1 (through $F^{\text{CIP}} = \frac{S_1 R^d}{R^f}$). Within a match involving a type-i client, the equilibrium traded notional is $X_i(\theta, \omega; S_1)$. Because a client trades only upon meeting a dealer, aggregate notional demand must account for both (i) the active mass of each type and (ii) the meeting probability $p_C(\theta)$. The aggregate net OTC notional demand contributed by type-i is $\mathcal{X}_i^{\text{OTC}}(\theta, \omega; S_1) \equiv U_i(\theta, \omega; S_1)\, p_C(\theta)\, X_i(\theta, \omega; S_1)$, so the total net OTC notional demand is the sum across types

$$\mathcal{X}^{\text{OTC}}(\theta, \omega; S_1) \equiv \sum_{i=1}^{I} \mathcal{X}_i^{\text{OTC}}(\theta, \omega; S_1), \tag{10}$$

Because X_i is signed, $\mathcal{X}^{\text{OTC}}$ is a *net* notional that can be positive or negative depending on the composition of types and on equilibrium trading directions.

Dealer inventory and hedging demand. Dealers intermediate client trades and therefore take the opposite side of client forward positions. Because dealers' aggregate forward payoff is linear in S_2, their net exposure to exchange-rate risk is proportional to the net notional $\mathcal{X}^{\text{OTC}}(\theta, \omega; S_1)$. We assume that, after the OTC round, dealers hedge this net FX exposure competitively in a Walrasian market by trading the foreign risk-free bond and do not take additional proprietary positions. Since one unit of the foreign bond yields R^f units of foreign currency at $t = 2$, full hedging requires $R^f\, H_D(\theta, \omega; S_1) = -\mathcal{X}^{\text{OTC}}(\theta, \omega; S_1)$, so dealers' hedging demand for the foreign bond is

$$H_D(\theta, \omega; S_1) = -\frac{\mathcal{X}^{\text{OTC}}(\theta, \omega; S_1)}{R^f}. \tag{11}$$

The next section formalizes the bond-market environment and closes the model by mapping dealer hedging demand $H_D(\theta, \omega; S_1)$ into equilibrium bond positions and, ultimately, into the spot exchange rate S_1 and the forward basis.

2.5 Walrasian Bond Market and Spot Determination

This section specifies the competitive bond market in which dealers hedge their OTC forward inventory. The spot exchange rate S_1 is pinned down by market clearing in the foreign bond, taking as inputs (i) the net dealer hedging demand implied by OTC trading, and (ii) residual bond demand from other domestic investors.

2.5.1 Domestic Spot Traders (Unhedged UIP Traders)

Domestic spot traders evaluate payoffs in domestic currency and can trade domestic and foreign risk-free bonds as well as the spot FX market. They implement a self-financed *unhedged* carry trade: borrow domestically at rate R^d, invest abroad at rate R^f, and bear FX risk. Preferences are mean–variance over terminal wealth.

Let H_d denote the trader's position in the foreign risk-free bond (in units of the foreign bond). At date 1, the trader borrows $S_1 H_d$ units of domestic currency, converts them into H_d units of foreign currency at the spot rate S_1, and purchases H_d units of the foreign bond. At date 2, the foreign position pays $R^f H_d$ in foreign currency, which the trader converts back into domestic currency at the spot rate S_2, yielding $S_2 R^f H_d$ in domestic currency. The trader then repays the domestic loan, $R^d S_1 H_d$, and earns profits $\Pi_d = S_2 R^f H_d - R^d S_1 H_d = Z H_d$, where the per-unit payoff is $Z = S_2 R^f - R^d S_1$.

Given mean–variance preferences, the trader's optimal foreign-bond demand is

$$H_d = \frac{\mathrm{E}[S_2] R^f - R^d S_1}{\gamma_d (R^f)^2 \mathrm{var}(S_2)}. \tag{12}$$

The expected per-unit payoff $\mathrm{E}[Z]$ measures the UIP deviation in domestic units. When foreign investment is attractive relative to UIP (high R^f and/or expected appreciation), $\mathrm{E}[Z] > 0$ and $H_d > 0$, so domestic spot traders go long the foreign bond, generating foreign-bond demand that depends negatively on S_1.

2.5.2 Inelastic Foreign-Bond Investor

A residual investor holds an exogenous position H_e in the foreign risk-free bond. This captures structural foreign-asset holdings driven by regulation, institutional mandates, or the economy's net foreign asset position. The purpose of this inelastic component is to ensure a well-defined residual demand in the spot FX determination.

2.5.3 Market Clearing and Spot Determination

Let $\overline{B}$ denote the (exogenous) supply of the foreign risk-free bond available to domestic investors. The Walrasian market clears at date 1, taking as given $\{R^d, R^f\}$ and the OTC-implied dealer hedging demand $H_D(\theta, \omega; S_1)$.

Market clearing in the foreign bond pins down the equilibrium spot exchange rate S_1:

$$H_e + H_d(S_1) + H_D(\theta, \omega; S_1) = \overline{B}.$$

Joint equilibrium (OTC–Walrasian fixed point). The general equilibrium jointly determines (θ, S_1) because (i) S_1 enters the OTC block through $F^{\text{CIP}} = \frac{S_1 R^d}{R^f}$ and hence affects within-match outcomes and entry values $\{J_i^U\}_{i \in \mathcal{I}}$, while (ii) θ affects dealers' aggregate hedging demand $H_D(\theta, \omega; S_1)$, which feeds back into the bond's market clearing condition. Formally, for a given state ω, an equilibrium is a pair (θ^*, S_1^*) such that

$$\theta^* = \frac{1}{V} \sum_{i=1}^{I} \overline{U}_i \, G\big(J_i^U(\theta^*, \omega; S_1^*)\big), \tag{13}$$

$$H_e + H_d(S_1^*) + H_D(\theta^*, \omega; S_1^*) = \overline{B}, \tag{14}$$

with $H_d(\cdot)$ given by Eq. (12) and $H_D(\cdot)$ given by Eq. (11) together with the aggregation in Eq. (10).

For expositional clarity, one can describe the solution as a two-step mapping: given a conjecture S_1, solve the OTC block for $\theta(S_1, \omega)$ (entry, composition, and aggregation), then impose Eq. (14) to update S_1; the equilibrium is the fixed point of this mapping. The next section characterizes the equilibrium, deriving conditions for existence and uniqueness, and exploring the conditions under which the feedback from OTC hedging can generate multiple equilibria and exchange rate instability.

3 Equilibrium Characterization and Quantitative Implications

In this section, we characterize the joint equilibrium of the OTC forward market and the Walrasian spot/bond market. We first establish the existence and potential uniqueness of the equilibrium, then analyze the conditions under which dealer hedging can generate multiple equilibria, and finally discuss the comparative statics and quantitative implications of the model.

3.1 Joint Equilibrium: Existence and Uniqueness

From the previous section, recall that an equilibrium is a pair (θ^*, S_1^*) satisfying the OTC entry condition Eq. (13) and the bond market clearing condition Eq. (14). The following proposition establishes conditions under which such an equilibrium exists and is unique.

Proposition 4 (Existence and uniqueness of joint equilibrium). *Fix ω and consider the discrete type space $\mathcal{I} = \{1, \ldots, I\}$. Define $\overline{F}^C \equiv \max_{i \in \mathcal{I}} F_i^C$ and*

$$\overline{S} \equiv \max\left\{0, \underbrace{\frac{R^f}{R^d}\overline{F}^C}_{\equiv \overline{S}^{\text{CIP}}}, \underbrace{\frac{\mathrm{E}[S_2]R^f - \gamma_d (R^f)^2 \mathrm{var}(S_2)\big(\overline{B} - H_e\big)}{R^d}}_{\equiv \overline{S}^{\text{Bd}}}\right\}.$$

Suppose that (i) Assumption 4 holds and $\sum_{i=1}^{I} \overline{U}_i < \infty$, so that by Proposition 3 for each $S_1 \in [0, \overline{S}]$ the OTC fixed point Eq. (13) admits a unique solution $\theta(S_1, \omega) \in [0, \bar{\theta}]$ (with $\bar{\theta} \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i$), and the function $D(\theta, S_1; \omega) \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i\, G\big(J_i^U(\theta, \omega; S_1)\big) - \theta$ is continuous on $[0, \bar{\theta}] \times [0, \overline{S}]$. Suppose further that (ii) there exists $\underline{S} \in [0, \overline{S}]$ such that the bond-market excess demand

$$\mathcal{Z}(S_1; \omega) \equiv H_e + H_d(S_1) + H_D(\theta(S_1, \omega), \omega; S_1) - \overline{B}$$

satisfies $\mathcal{Z}(\underline{S}; \omega) \geq 0$. Then there exists at least one joint equilibrium (θ^*, S_1^*) satisfying Eqs. (13) and (14).

Moreover, if $\mathcal{Z}(\cdot; \omega)$ is strictly decreasing on $[\underline{S}, \overline{S}]$—for example, if the Walrasian bond market is sufficiently deep in the sense that

$$\sup_{S_1 \in [\underline{S}, \overline{S}]} \frac{d}{dS_1} H_D(\theta(S_1, \omega), \omega; S_1) < -\frac{d}{dS_1} H_d(S_1) = \frac{R^d}{\gamma_d (R^f)^2 \mathrm{var}(S_2)},$$

then the joint equilibrium is unique.

Proof. See Appendix A.3.4. □

Proposition 4 defines bounds $[\underline{S}, \overline{S}]$ for the spot exchange rate to guarantee the existence of an equilibrium. The upper bound $\overline{S}$ guarantees negative excess demand when the spot rate is sufficiently high. The first element in its definition, $\overline{S}^{\text{CIP}}$, corresponds to the CIP evaluated at the largest reservation rate across clients. This guarantees a positive demand for forwards and a negative dealer position in the bond market ($H_D < 0$). The second component, $\overline{S}^{\text{Bd}}$, defines a minimum value for the spot rate such that residual bond excess demand (net of dealers) is negative. Since $\overline{S}$ is the maximum of these elements, $\mathcal{Z}(\overline{S}; \omega) \leq 0$ holds.

Condition (ii), in turn, is a mild sign restriction that ensures a crossing of bond excess demand on $[\underline{S}, \overline{S}]$. Economically, it requires that at some sufficiently low spot rate $\underline{S}$, residual foreign-bond demand (from unhedged UIP traders and dealers' hedging needs) exceeds effective supply $\overline{B} - H_e$, so that market clearing requires a higher S_1.

3.2 Dealer Hedging Feedback and Multiplicity

Since the OTC block delivers a single-valued map $S_1 \mapsto \theta(S_1, \omega)$, the joint equilibrium reduces to solving the scalar equation $\mathcal{Z}(S_1; \omega) = 0$. Because $H_d(S_1)$ is affine and strictly decreasing, multiple equilibria require that the dealer-hedging term $S_1 \mapsto H_D(\theta(S_1, \omega), \omega; S_1)$ be sufficiently increasing (or non-monotone) over some range so as to offset the UIP-trader demand.

A fixed-point representation. Using Eq. (12), bond-market clearing Eq. (14) can be written as a fixed point in the spot rate:

$$S_1 = \underbrace{\frac{\mathrm{E}[S_2]R^f - \gamma_d (R^f)^2 \mathrm{var}(S_2)\left(\overline{B} - H_e\right)}{R^d}}_{\equiv S_1^{\mathrm{UIP}}} + \underbrace{\frac{\gamma_d (R^f)^2 \mathrm{var}(S_2)}{R^d}}_{\equiv \chi_d} H_D(\theta(S_1,\omega),\omega;S_1). \tag{15}$$

Eq. (15) makes clear that dealer hedging demand is the only source of nonlinearity in the Walrasian block. If H_D were exogenous or constant in S_1, the spot rate would be pinned down uniquely by S_1^{UIP}. With endogenous dealer hedging, a higher (more positive) H_D increases spot demand for the foreign bond and therefore requires a higher S_1 to clear the market, which we study next.

Local uniqueness and the role of dealer hedging slope. Differentiating $\mathcal{Z}(S_1;\omega) = 0$ yields

$$\mathcal{Z}'(S_1;\omega) = \frac{d}{dS_1}H_d(S_1) + \frac{d}{dS_1}H_D(\theta(S_1,\omega),\omega;S_1) = -\frac{R^d}{\gamma_d (R^f)^2 \mathrm{var}(S_2)} + \frac{d}{dS_1}H_D(\theta(S_1,\omega),\omega;S_1). \tag{16}$$

Thus, $\mathcal{Z}(\cdot;\omega)$ is strictly decreasing (and the joint equilibrium is unique) if dealer hedging is not "too increasing" in S_1, i.e. if $\frac{d}{dS_1}H_D < \frac{R^d}{\gamma_d (R^f)^2 \mathrm{var}(S_2)}$. Conversely, any multiplicity must come from a strong positive slope of the equilibrium dealer-hedging schedule.

Geometry and stability. Define the bond-market implied mapping $\Phi(S_1;\omega) \equiv S_1^{\mathrm{UIP}} + \chi_d H_D(\theta(S_1,\omega),\omega;S_1)$. Joint equilibria are fixed points $S_1 = \Phi(S_1;\omega)$. The local slope satisfies $\Phi'(S_1;\omega) = \chi_d \frac{d}{dS_1}H_D(\theta(S_1,\omega),\omega;S_1)$, so a sufficient condition for uniqueness on the relevant interval is $\Phi'(S_1;\omega) < 1$ on $[\underline{S},\overline{S}]$. If $\Phi(\cdot;\omega)$ is S-shaped and crosses the $45°$ line three times, the middle fixed point is unstable under standard tatonnement, while the two outer fixed points are stable.

Corollary 2 (Sufficient uniqueness via bond-market slope). *Fix ω and suppose the conditions of Proposition 4 hold. If $\Phi(\cdot;\omega)$ is differentiable on $[\underline{S},\overline{S}]$ and satisfies $\sup_{S_1\in[\underline{S},\overline{S}]}\Phi'(S_1;\omega) < 1$, then the joint equilibrium is unique.*

Proof. Let $f(S_1;\omega) \equiv \Phi(S_1;\omega) - S_1$. Then $f'(S_1;\omega) = \Phi'(S_1;\omega) - 1 < 0$ for all $S_1 \in [\underline{S},\overline{S}]$, so $f(\cdot;\omega)$ is strictly decreasing on that interval and can cross zero at most once. Since Proposition 4 delivers existence of a root in $[\underline{S},\overline{S}]$, it must be unique. □

Why dealer hedging can be non-monotone in S_1. Recall that $H_D(\theta,\omega;S_1) = -\mathcal{X}^{\mathrm{OTC}}(\theta,\omega;S_1)/R^f$, where $\mathcal{X}^{\mathrm{OTC}}$ is net OTC notional. Along the equilibrium tightness schedule,

$$\frac{d}{dS_1}H_D(\theta(S_1,\omega),\omega;S_1) = \underbrace{\frac{\partial H_D(\theta,\omega;S_1)}{\partial S_1}}_{\text{direct CIP effect}} + \underbrace{\frac{\partial H_D(\theta,\omega;S_1)}{\partial \theta}\theta'(S_1,\omega)}_{\text{indirect congestion effect}}, \qquad \theta = \theta(S_1,\omega). \tag{17}$$

The first term is a *direct* CIP channel: higher S_1 raises $F^{\mathrm{CIP}} = \frac{S_1 R^d}{R^f}$ and therefore shifts within-match quantities and match surpluses. In our linear–quadratic environment, holding θ fixed this direct effect makes dealer hedging *decrease* in S_1. The second term is an *indirect* congestion channel: a change in S_1 reshuffles which client types participate and hence changes tightness, which in turn affects meeting probabilities, bargaining weights, and dealer costs.

3.2.1 The Direct CIP Channel

In the within-match equilibrium of Corollary 1, each type-i traded notional is linear in the mispricing gap $m_i(S_1) \equiv F^{\mathrm{CIP}}(S_1) - F_i^C$, where $F^{\mathrm{CIP}}(S_1) = \frac{S_1 R^d}{R^f}$ and $F_i^C = \mathrm{E}[S_2] - \gamma_i \mathrm{var}(S_2)\widetilde{X}_i R^f$. In particular,

$$X_i(\theta,\omega;S_1) = A_i(\theta,\omega)\, m_i(S_1), \qquad A_i(\theta,\omega) \equiv \frac{\alpha^{(p)}(\theta)}{\gamma_i \mathrm{var}(S_2) + \frac{\alpha^{(p)}(\theta)\kappa(\theta,\omega)}{2(R^f)^2}} > 0, \tag{18}$$

$$J_i^U(\theta,\omega;S_1) = B_i(\theta,\omega)\, m_i(S_1)^2, \qquad B_i(\theta,\omega) \equiv \frac{\delta_C}{1-\delta_C} p_C(\theta)\, \frac{\alpha^{(p)}(\theta)^2\, \gamma_i \mathrm{var}(S_2)}{\Big(\gamma_i \mathrm{var}(S_2) + \frac{\alpha^{(p)}(\theta)\kappa(\theta,\omega)}{2(R^f)^2}\Big)^2} > 0. \tag{19}$$

Assume for this local analysis that G is continuously differentiable with density $g(\cdot) \equiv G'(\cdot) \geq 0$. Holding (θ,ω) fixed and using Eqs. (18) and (19), we obtain

$$\frac{\partial H_D(\theta,\omega;S_1)}{\partial S_1} = -\frac{p_C(\theta)}{R^f}\frac{R^d}{R^f}\sum_{i=1}^{I} \overline{U}_i A_i(\theta,\omega)\Big(G\big(J_i^U(\theta,\omega;S_1)\big) + 2J_i^U(\theta,\omega;S_1)g\big(J_i^U(\theta,\omega;S_1)\big)\Big) < 0, \tag{20}$$

so the direct effect of a higher spot rate is always to increase net OTC notional and therefore make dealer hedging demand more negative (more foreign-bond supply by dealers). Holding θ fixed, a higher S_1 raises $F^{\mathrm{CIP}} = \frac{S_1 R^d}{R^f}$ and shifts the mispricing gaps $m_i(S_1)$ upward; in this linear–quadratic environment this increases traded notionals and participation, raising $\mathcal{X}^{\mathrm{OTC}}$ and therefore lowering $H_D = -\mathcal{X}^{\mathrm{OTC}}/R^f$.[5] Any local increase of H_D in S_1 must therefore be driven by the endogenous response of tightness in the second term of Eq. (17).

3.2.2 The Congestion Channel

If the entry mapping is differentiable, the implicit function theorem delivers the following expression for tightness elasticity

$$\theta'(S_1,\omega) = \left.\frac{\frac{\partial T(\theta,\omega;S_1)}{\partial S_1}}{1 - \frac{\partial T(\theta,\omega;S_1)}{\partial \theta}}\right|_{\theta=\theta(S_1,\omega)}, \qquad T(\theta,\omega;S_1) \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i G\big(J_i^U(\theta,\omega;S_1)\big). \tag{21}$$

[5] Eq. (20) characterizes local slopes under differentiability of G. With non-smooth participation distributions (kinks or mass points), the direct CIP effect remains weakly negative and the congestion term can be characterized using one-sided derivatives or subgradients.

Because $\partial T/\partial\theta \le 0$ under Assumption 4, the denominator in Eq. (21) is at least one, so the sign and magnitude of $\theta'(S_1,\omega)$ are governed by $\partial T/\partial S_1$. Under the same differentiability condition as above, we have

$$\frac{\partial T(\theta,\omega;S_1)}{\partial S_1} = \frac{1}{V}\sum_{i=1}^{I}\overline{U}_i g\big(J_i^U(\theta,\omega;S_1)\big)\frac{\partial J_i^U(\theta,\omega;S_1)}{\partial S_1}, \tag{22}$$

$$\frac{\partial T(\theta,\omega;S_1)}{\partial \theta} = \frac{1}{V}\sum_{i=1}^{I}\overline{U}_i g\big(J_i^U(\theta,\omega;S_1)\big)\frac{\partial J_i^U(\theta,\omega;S_1)}{\partial \theta} \le 0. \tag{23}$$

In the mean–variance model, $J_i^U(\theta,\omega;S_1)$ depends on S_1 only through the *mispricing* gap $m_i(S_1) \equiv F^{\text{CIP}}(S_1) - F_i^C$ and takes the quadratic form in Eq. (19), so $\partial_{S_1} J_i^U(\theta,\omega;S_1)$ is proportional to $m_i(S_1)$. Hence, types with $m_i(S_1) > 0$ tend to increase participation as S_1 rises, while types with $m_i(S_1) < 0$ tend to reduce participation as S_1 rises. Importantly, the density $g\big(J_i^U\big)$ weights which types are close to the participation margin and therefore dominate the response of tightness to the spot rate.

Combining Eqs. (17) and (21), dealer hedging responds to the spot rate according to

$$\frac{d}{dS_1}H_D(\theta(S_1,\omega),\omega;S_1) = \partial_{S_1}H_D(\theta,\omega;S_1) + \partial_\theta H_D(\theta,\omega;S_1)\frac{\partial_{S_1}T(\theta,\omega;S_1)}{1-\partial_\theta T(\theta,\omega;S_1)}, \qquad \theta = \theta(S_1,\omega), \tag{24}$$

which makes transparent the feedback loop from spot to OTC entry to dealer hedging.

Economic conditions for multiplicity. Equation Eq. (16) shows that multiple joint equilibria require $\mathcal{Z}(\cdot;\omega)$ to fail strict monotonicity, which can happen only if the indirect term in Eq. (17) is strong enough to overturn the negative direct CIP effect. Multiplicity is therefore more likely when (i) the bond market is *shallow* (high γ_d and/or high $\text{var}(S_2)$, so that $-H_d'(S_1)$ is small), (ii) participation is highly elastic (high density of G around the relevant entry values, so that $\partial_{S_1}T$ is large), and (iii) congestion strongly affects within-match terms and dealer costs (large $|p_C'(\theta)|$, strongly decreasing $\alpha^{(p)}(\theta)$, and strongly increasing $\kappa(\theta,\omega)$).

The client-side fundamentals $(\gamma_i, \widetilde{X}_i)$ enter through the reservation rates $\{F_i^C\}$ and therefore through the cross-sectional distribution of mispricing gaps $\{m_i(S_1)\}$, which governs the sign and size of $\partial_{S_1}T$ in Eq. (22) and hence the strength of the congestion feedback in Eq. (24). Stress ω affects multiplicity through $\kappa(\theta,\omega)$: higher ω raises balance-sheet curvature, reducing $A_i(\theta,\omega)$ and $B_i(\theta,\omega)$ and therefore damping both within-match volumes and entry values, which ceteris paribus flattens the dealer-hedging schedule; the net effect also depends on how stress reshapes participation elasticities (the density g around the entry margin).

In the absence of any of these channels (e.g., if all clients participate so θ is fixed, or if congestion does not affect within-match terms), the dealer-hedging schedule is monotone decreasing and the joint equilibrium is unique.

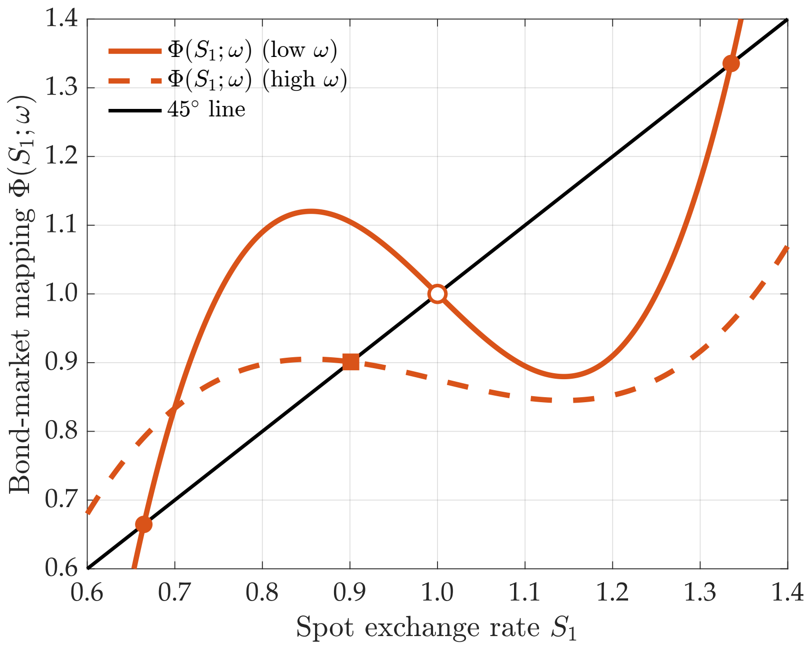


Figure 3. Equilibrium determination in the spot market

Notes: The figure plots the mapping $\Phi(S_1;\omega) \equiv S_1^{\mathrm{UIP}} + \chi_d H_D(\theta(S_1,\omega),\omega;S_1)$ against the 45-degree line. Fixed points correspond to equilibrium spot rates S_1^*. Under low financial stress (solid curve), the mapping can be S-shaped, generating three intersections (two stable, one unstable). Under high financial stress (dashed curve), increased balance-sheet curvature flattens the mapping, restoring uniqueness. The transition from multiple to unique equilibria illustrates how dealer balance-sheet costs discipline exchange rate dynamics.

Fig. 3 provides a graphical representation of equilibrium determination in the spot market. Recall from Eq. (15) that the spot rate must satisfy the fixed-point condition $S_1 = \Phi(S_1;\omega)$, where $\Phi(S_1;\omega) \equiv S_1^{\mathrm{UIP}} + \chi_d H_D(\theta(S_1,\omega),\omega;S_1)$ captures the combined effect of UIP traders and dealer hedging demand. The 45-degree line represents points where $S_1 = \Phi(S_1;\omega)$; intersections are equilibria. We illustrate the point considering changes in financial stress ω.

Under low financial stress (solid curve), the mapping Φ can become S-shaped. This occurs when the congestion channel in Eq. (24) is strong enough to make dealer hedging demand locally increasing in S_1. The S-shape generates three intersections with the 45-degree line: two stable equilibria (the outer crossings) and one unstable equilibrium (the middle crossing). This configuration illustrates how identical fundamentals can support multiple exchange rate realizations depending on market expectations and the coordination of dealer and client behavior.

Under high financial stress (dashed curve), increased balance-sheet curvature $\kappa(\theta,\omega)$ dampens both the direct and indirect channels in Eq. (24). The mapping flattens, eliminating the S-shape and restoring a unique equilibrium. The figure thus captures a central message of the paper: financial stress, by tightening dealer balance-sheet constraints, can paradoxically stabilize the exchange rate by eliminating multiplicity, even as it widens the CIP basis. This insight connects our model to policy discussions about the trade-offs between market liquidity and stability in emerging economies.

4 Conclusion

This paper develops a tractable model of joint spot and forward exchange rate determination in an environment where spot FX and bonds trade competitively, while FX forwards are negotiated over the counter under search frictions and bilateral bargaining. On the demand side, a general class of mean–variance clients choose signed forward notionals, allowing the model to accommodate both natural hedgers with exogenous currency exposures and investors with speculative demand. On the supply side, dealers intermediate client trades, hedge net forward inventory in the competitive bond market, and face convex balance-sheet costs whose curvature increases with market tightness and a financial stress state.

The model delivers three main messages. First, OTC frictions endogenize forward pricing and quantities through a congestion channel. Higher tightness worsens meeting prospects and shifts bargaining power toward dealers, compressing within-match volumes and moving negotiated quotes toward clients' reservation rates (i.e., away from dealers' break-even rates). This generates an endogenous forward basis relative to the covered interest parity benchmark. Second, balance-sheet scarcity, captured by higher curvature under stress, disciplines intermediation. It reduces traded volumes and entry values, widens the basis, and reshapes the distribution of equilibrium notionals across client types. Third, because dealers hedge aggregate OTC inventory in the competitive bond market, OTC activity feeds back into the spot rate. The resulting joint equilibrium can be characterized as a scalar clearing condition in the spot rate, with existence under mild sign restrictions and uniqueness whenever dealer hedging is not too increasing in the spot rate. Importantly, multiplicity can arise only if the congestion feedback is strong enough to overturn the negative direct effect of covered interest parity on dealer hedging demand. This requires a combination of shallow bond markets, elastic participation around the entry margin, and strong sensitivity of within-match terms to tightness and stress.

Next steps are to (i) derive comparative statics for the forward basis, traded volumes, and the spot rate with respect to tightness and stress; (ii) map the model to data by identifying empirical proxies for tightness and stress and matching moments for forward quotes and quantities; and (iii) quantify welfare and policy implications of changes in intermediation capacity and market structure, including interventions that relax dealer balance-sheet constraints during stress episodes.

References

Alfaro, L., Calani, M., and Varela, L. (2023). Granular Corporate Hedging Under Dominant Currency. Working Paper 28910, National Bureau of Economic Research. 6

Atkenson, A., Eisfeldt, A., and Weil, P. (2015). Entry and Exit in OTC Derivatives Markets. *Econometrica*, 83(6):2231–2292. 3

Avdjiev, S., Du, W., Koch, C., and Shin, H. (2019). The Dollar, Bank Leverage, and Deviations from Covered Interest Parity. *American Economic Review: Insights*, 1(2):193–208. 3

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The Nash Bargaining Solution in Economic Modelling. *The RAND Journal of Economics*, 17(2):176–188. 10

Borio, C., McCauley, R., McGuire, P., and Sushko, V. (2016). Covered Interest Parity Lost: Understanding the Cross-Currency Basis. *BIS Quarterly Review*, pages 45–64. 3

Chen, N. and Zhou, H. (2025). Managing Emerging Market Currency Risk. Technical report. 5

De Leo, P., Keller, L., and Zou, D. (2025). Speculation, Forward Exchange Demand, and CIP Deviations in Emerging Economies. Technical report. 5, 6

den Haan, W., Ramey, G., and Watson, J. (2000). Job Destruction and Propagation of Shocks. *American Economic Review*, 90(3):482–498. 13

Du, W., Tepper, A., and Verdelhan, A. (2018). Deviations from Covered Interest Rate Parity. *Journal of Finance*, 73(3):915–957. 3

Duffie, D., Gârleanu, N., and Pedersen, L. (2005). Over-the-Counter Markets. *Econometrica*, 73(6):1815–1847. 3

Evans, M. and Lyons, R. (2002). Order Flow and Exchange Rate Dynamics. *Journal of Political Economy*, 110(1):180–180. 3

Gabaix, X. and Maggiori, M. (2015). International Liquidity and Exchange Rate Dynamics. *The Quarterly Journal of Economics*, 130(3):1369–1420. 3

Huang, W., Ranaldo, A., Schrimpf, A., and Somogyi, F. (2025). Constrained Liquidity Provision in Currency Markets. *Journal of Financial Economics*, 167:1–19. 3

Lagos, R. and Rocheteau, G. (2009). Liquidity in Asset Markets with Search Frictions. *Econometrica*, 77(2):403–426. 3

Lagos, R., Rocheteau, G., and Weill, P. (2011). Crises and Liquidity in Over-The-Counter Markets. *Journal of Economic Theory*, 146(6):2169–2205. 3

Liao, G. and Zhang, T. (2025). The Hedging Channel of Exchange Rate Determination. *The Review of Financial Studies*, 38(1):1–38. 5

Rime, D., Schrimpf, A., and Syrstad, O. (2022). Covered Interest Parity Arbitrage . *The Review of Financial Studies*, 35(11):5185–5227. 3

Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50(1):97–109. 3, 10

A Theoretical Results

A.1 Dealer Balance-Sheet Curvature κ as Function of Market Conditions

Here we discuss some alternative motivations of why the dealer-specific curvature cost may depend on market-wide conditions.

Equilibrium pricing of scarce funding/collateral (Macro-finance view). A dealer's intermediation capacity relies on inputs such as secured funding, margin/collateral, regulatory capital, and internal risk budgets. The effective price of these inputs is not fixed; it responds to aggregate demand for dealer balance sheet. When tightness θ rises (more clients per dealer), aggregate demand for intermediation per dealer increases, and funding/collateral markets and internal constraints become tighter. This raises the marginal cost of expanding positions and can steepen the dealer's cost schedule.

This logic is captured by allowing $\frac{\partial\kappa(\theta,\omega)}{\partial\theta} > 0$ and $\frac{\partial\kappa(\theta,\omega)}{\partial\omega} > 0$. The former reflects the equilibrium impact of congestion on the marginal convexity of balance-sheet usage, while the latter reflects that stress increases the shadow cost of balance sheet.

Some closed-form examples are

$$\kappa(\theta,\omega) = \kappa_0 \exp(\eta_\theta \theta)\exp(\eta_\omega \omega), \qquad \kappa_0 > 0,\ \eta_\theta > 0,\ \eta_\omega > 0,$$
$$\kappa(\theta,\omega) = \kappa_0\big(1+\eta_\omega\omega\big)\big(1+\eta_\theta\theta^2\big), \qquad \eta_\theta > 0,\ \eta_\omega > 0,$$

which guarantee positivity and satisfy the sign restrictions.

Mean-field utilization of dealer balance sheet (OTC micro view). One may object that congestion is dealer-specific, so that costs should depend on a "micro tightness" θ_i faced by dealer i. In a large symmetric OTC market with many ex ante identical dealers, however, a standard mean-field logic applies: each dealer's *expected* flow of client contacts and trades is governed by aggregate tightness θ through the dealer matching probability $p_D(\theta)$. Hence, in a stationary equilibrium, dealer utilization is approximately deterministic and identical across dealers, and can be written as a function of θ alone.

To see this point more clearly, let B denote a dealer's total balance-sheet usage within the OTC segment (e.g., a summary statistic for outstanding balance-sheet-intensive intermediation). Suppose the primitive cost of balance-sheet usage is convex and depends on the stress state:

$$C(B;\omega) = \frac{\kappa_0(\omega)}{2}B^2, \qquad \kappa_0(\omega) > 0, \quad \kappa_0'(\omega) > 0.$$

In a stationary equilibrium, the dealer operates around a baseline utilization $\bar{B}(\theta)$ that is increasing in tightness because the dealer matches more often when there are more clients per dealer (formally, $p'_D(\theta) > 0$):

$$\bar{B}'(\theta) > 0.$$

Consider the incremental cost of accommodating an additional position of size K around the baseline $\bar{B}(\theta)$:

$$\begin{aligned} C(\bar{B}(\theta) + K; \omega) - C(\bar{B}(\theta); \omega) &= \frac{\kappa_0(\omega)}{2}\big(\bar{B}(\theta) + K\big)^2 - \frac{\kappa_0(\omega)}{2}\bar{B}(\theta)^2 \\ &= \kappa_0(\omega)\bar{B}(\theta)\,K + \frac{\kappa_0(\omega)}{2}K^2. \end{aligned}$$

This expression has two components: (i) a *linear* term proportional to baseline utilization $\bar{B}(\theta)$, capturing the opportunity cost of scarce balance sheet when the dealer is already busy; and (ii) a *quadratic* term capturing intrinsic convexity. Both terms become larger when tightness increases (through $\bar{B}(\theta)$) and when stress increases (through $\kappa_0(\omega)$). A reduced-form representation that preserves tractability is therefore to summarize the net incremental cost around equilibrium utilization using an effective curvature $\kappa(\theta, \omega)$ that is increasing in θ and ω.

A particularly transparent reduced form is to let κ depend on θ via $p_D(\theta)$:

$$\kappa(\theta, \omega) = \kappa_0(\omega)\,\phi\big(p_D(\theta)\big), \qquad \phi' > 0,$$

since $p_D(\theta)$ is the primitive object that measures how frequently a dealer is contacted/matched and thus how congested her desk is in expectation.

Dealer attention/risk-budget scarcity (opportunity cost view). In OTC markets, intermediation uses scarce resources beyond funding: quoting capacity, risk-manager attention, operational bandwidth, and internal risk-budget headroom. When θ is high, each dealer faces more requests per unit time. Even if the dealer is atomistic and price-taking, the opportunity cost of allocating balance sheet to any given client increases because the dealer expects many alternative requests. This can be summarized as an increase in the marginal convexity of scaling up positions, motivating $\frac{\partial \kappa(\theta,\omega)}{\partial \theta} > 0$.

A.2 A Microfoundation Linking Breakdown Probabilities to Matching Probabilities

In the baseline search block, market tightness θ determines meeting probabilities for unmatched agents, $p_C(\theta)$ and $p_D(\theta)$. In the bargaining block, we introduce within-match breakdown probabilities after a rejected offer, $\lambda_C(\theta)$ and $\lambda_D(\theta)$. Because both sets of objects move with tightness, a natural question is whether the model is over-parameterized. This appendix provides a clean microfoundation in which within-match breakdown hazards are endogenously linked to the same primitives that govern unmatched matching, while preserving the economic distinction between the extensive (matching) and intensive (within-match) margins.

Two distinct margins. The meeting probabilities $p_C(\theta)$ and $p_D(\theta)$ govern the probability of forming a match when an agent is currently unmatched. By contrast, $\lambda_C(\theta)$ and $\lambda_D(\theta)$ govern the probability that an *ongoing negotiation* breaks down after a rejection, because the rejected proposer activates an outside option (shop-around, re-quoting, reallocating scarce capacity, or timing out). Thus, even if both objects depend on θ, they affect different parts of the game: $p_i(\theta)$ shapes *time-to-match* and volumes, while $\lambda_i(\theta)$ shapes *within-match continuation values* and hence equilibrium bargaining weights and prices conditional on trade.

A.2.1 A Simple Timing Protocol for Negotiations and Outside Opportunities

Within-match time and alternative opportunities. Consider a matched client–dealer pair that is bargaining in discrete rounds $n = 0, 1, 2, \ldots$. Interpret a "bargaining round" as a short time interval of length $\Delta > 0$ (e.g., minutes or hours within a trading day) during which (i) one side posts a quote, (ii) the other side responds, and (iii) if the quote is rejected, the proposer considers abandoning the negotiation.

During the interval following a rejection, the rejected proposer may receive an *outside opportunity* to restart bargaining elsewhere (shop-around) or to reallocate capacity to another counterparty. Model the arrival of such outside opportunities as a Poisson process.

Poisson arrivals. Let $\mu_i(\theta) \geq 0$ denote the arrival intensity of outside opportunities for side $i \in \{C, D\}$ during negotiation. Conditional on a rejection, the probability that at least one outside opportunity arrives within the next Δ units of time is

$$\lambda_i(\theta) = 1 - \exp\big(-\mu_i(\theta)\Delta\big), \qquad i \in \{C, D\}. \tag{A.1}$$

We interpret $\lambda_i(\theta)$ as the probability that the negotiation breaks down after a rejection because the rejected proposer finds it optimal (or is forced) to terminate the current interaction before the next bargaining round.

A.2.2 Linking Arrival Intensity to Unmatched Matching Probabilities

A parsimonious link is to assume that the intensity of outside opportunities during negotiation is proportional to the unmatched meeting probability implied by the matching technology.

Formally, let

$$\mu_C(\theta) = \nu_C \, p_C(\theta), \tag{A.2}$$

$$\mu_D(\theta) = \nu_D \, p_D(\theta), \tag{A.3}$$

where $\nu_C, \nu_D > 0$ are shifters capturing how quickly each side can act on opportunities (e.g., operational capacity, urgency, or propensity to shop around). Combining Eqs. (A.1) to (A.3) yields the closed-form mapping

$$\lambda_C(\theta) = 1 - \exp\big(-\nu_C p_C(\theta)\Delta\big), \tag{A.4}$$

$$\lambda_D(\theta) = 1 - \exp\big(-\nu_D p_D(\theta)\Delta\big). \tag{A.5}$$

Since only the product $\nu_i\Delta$ matters, one can absorb Δ into ν_i and write $\lambda_i(\theta) = 1 - \exp\big(-\nu_i p_i(\theta)\big)$ without loss of generality.

The meeting probabilities $p_i(\theta)$ summarize how easy it is to contact alternative counterparties in the market. Eq. (A.4) indicates that when clients face a low meeting probability (high tightness), they are less likely to abandon a negotiation after being rejected because alternative dealers are harder to reach within the relevant time window. On the other hand, Eq. (A.5) indicates that when dealers face a high meeting probability (high tightness), they are more likely to abandon a negotiation after being rejected because alternative clients are plentiful.

A.2.3 Implications for signs and comparative statics

Monotonicity in tightness. Under standard matching assumptions, $p'_C(\theta) < 0$ and $p'_D(\theta) > 0$. From Eq. (A.4) and Eq. (A.5), we obtain

$$\frac{d\lambda_C(\theta)}{d\theta} = \exp\big(-\nu_C p_C(\theta)\Delta\big)\,\nu_C\Delta\, p'_C(\theta) < 0, \tag{A.6}$$

$$\frac{d\lambda_D(\theta)}{d\theta} = \exp\big(-\nu_D p_D(\theta)\Delta\big)\,\nu_D\Delta\, p'_D(\theta) > 0. \tag{A.7}$$

Hence the asymmetric-breakdown sign pattern $\lambda'_C(\theta) < 0$ and $\lambda'_D(\theta) > 0$ emerges endogenously from the same primitive forces that govern unmatched matching.

Identification and interpretation. Even under Eqs. (A.4) to (A.5), the model is not redundant because $p_i(\theta)$ continues to determine the extensive margin (how often agents meet when unmatched), while $\lambda_i(\theta)$ determines the intensive margin of bargaining (how likely negotiations are to continue after disagreements). The mapping disciplines the direction and curvature of $\lambda_i(\theta)$ using primitives already present in the search block, reducing concerns about over-parameterization and strengthening the economic interpretation of the asymmetric breakdown assumptions.

Alternative reduced forms. While Eq. (A.1) is convenient and grounded in a standard arrival-time argument, any increasing function $\Lambda(\cdot)$ with $\Lambda(0) = 0$ could be used to map meeting probabilities into breakdown probabilities, e.g., $\lambda_i(\theta) = \Lambda_i(\nu_i p_i(\theta))$ with $\Lambda_i' > 0$. The exponential form above is attractive because it is bounded in $[0, 1)$ and preserves analytical tractability.

A.3 Proofs

A.3.1 Proof of Proposition 1

Proof. We proceed in steps and solve for equilibrium surplus splits. We consider the case where the client proposes first. Throughout, all continuation values and surpluses are defined in incremental terms, net of the outside option of returning to the unmatched market state. Hence, if bargaining breaks down, the continuation payoff in the surplus-splitting problem is normalized to zero.

Step 1 (define within-round equilibrium surpluses). Let Σ_C^C denote the equilibrium surplus obtained by the client when it is the client's turn to propose. Let Σ_C^D denote the equilibrium surplus obtained by the client when it is the dealer's turn to propose. Define analogously Σ_D^C, Σ_D^D for the dealer. Because the pie is S in any agreement,

$$\Sigma_C^C + \Sigma_D^C = S, \qquad \Sigma_C^D + \Sigma_D^D = S.$$

Step 2 (acceptance constraints from one-shot deviations). In a client-proposal round, the dealer (as the responder) will accept any offer that gives her at least her discounted expected continuation value from rejecting. If she rejects, the client (whose offer was rejected) may walk away with probability $\lambda_C(\theta)$, so with probability $q_C(\theta)$ bargaining continues to the next round (where the dealer proposes). Hence the dealer's continuation value from rejecting equals $\delta_D q_C(\theta)\Sigma_D^D$. In equilibrium the client sets the offer to make the dealer indifferent

$$\Sigma_D^C = \delta_D q_C(\theta)\Sigma_D^D. \tag{A.8}$$

Similarly, in a dealer-proposal round, if the client rejects then the dealer (whose offer was rejected) may walk away with probability $\lambda_D(\theta)$, so with probability $q_D(\theta)$ bargaining continues to the next round (where the client proposes). Thus the client's continuation value from rejecting equals $\delta_C q_D(\theta)\Sigma_C^C$, and in equilibrium the dealer makes the client indifferent

$$\Sigma_C^D = \delta_C q_D(\theta)\Sigma_C^C. \tag{A.9}$$

Step 3 (solve the system). Using $\Sigma_D^C = S - \Sigma_C^C$ and $\Sigma_D^D = S - \Sigma_C^D$ in Eq. (A.8)

$$S - \Sigma_C^C = \delta_D q_C(\theta)\big(S - \Sigma_C^D\big). \tag{A.10}$$

Substituting Eq. (A.9) into Eq. (A.10) and rearranging, we get

$$S\big(1 - \delta_D q_C(\theta)\big) = \Sigma_C^C\big(1 - \delta_C\delta_D q_C(\theta)q_D(\theta)\big).$$

Therefore,

$$\Sigma_C^C = S\,\frac{1 - \delta_D q_C(\theta)}{1 - \delta_C\delta_D q_C(\theta)q_D(\theta)}.$$

Since in the client-first protocol agreement occurs immediately at $n = 0$ (the client-proposal round), the client's equilibrium share is

$$\alpha^{CF}(\theta) \equiv \frac{\Sigma_C^C}{S} = \frac{1 - \delta_D q_C(\theta)}{1 - \delta_C\delta_D q_C(\theta)q_D(\theta)}.$$

Step 4 (map the surplus share into the equilibrium forward). In any agreement, the client's realized surplus is $X(F - F^C)$. By definition of α^{CF},

$$X\big(F^{CF}(\theta) - F^C\big) = \alpha^{CF}(\theta)\,X(F^D - F^C).$$

Divide by X and rearrange to obtain

$$F^{CF}(\theta) = (1 - \alpha^{CF}(\theta))F^C + \alpha^{CF}(\theta)F^D.$$

Finally, $\alpha^{CF}(\theta) \in (0,1)$ follows from $\delta_D q_C(\theta) \in (0,1)$ and $\delta_C\delta_D q_C(\theta)q_D(\theta) \in (0,1)$.

Now we focus on the case where the dealer offers first. From our previous derivations, we know the client surplus when the client proposes, Σ_C^C. Using Eq. (A.9), we know that $\Sigma_C^D = \delta_C q_D(\theta)\Sigma_C^C$. In the dealer-first protocol, agreement occurs immediately in the dealer-proposal round ($n = 0$), so

$$\alpha^{DF}(\theta) \equiv \frac{\Sigma_C^D}{S} = \delta_C q_D(\theta) \frac{\Sigma_C^C}{S} = \delta_C q_D(\theta)\, \alpha^{CF}(\theta),$$

which yields the closed-form expression stated in the proposition. The expression for $F^{DF}(\theta)$ follows from the same linear-surplus mapping as in Step 4, while $\alpha^{DF}(\theta) \in (0,1)$ follows from $\delta_C q_D(\theta) \in (0,1)$ and $\delta_C \delta_D q_C(\theta) q_D(\theta) \in (0,1)$.

Finally, note that the forward rate agreed by a matched client–dealer pair lies between their reservation rates and is a weighted average of F^C and F^D, with the weights determined endogenously by discount factors, market tightness (through $\lambda_C(\theta)$ and $\lambda_D(\theta)$), and the bargaining protocol. Since $0 < \alpha^{(p)}(\theta) < 1$, $F(\theta)$ is a strict convex combination of F^C and F^D and therefore lies strictly between them: $\min\{F^C, F^D\} < F(\theta) < \max\{F^C, F^D\}$ whenever $F^C \neq F^D$. □

A.3.2 Proof of Proposition 2

Proof. First, let us write α^{CF} as a function of (q_C, q_D) as follows

$$\alpha^{CF}(q_C, q_D) = \frac{1 - \delta_D q_C}{1 - \delta_C \delta_D q_C q_D} \equiv \frac{N}{\Delta}, \qquad N \equiv 1 - \delta_D q_C, \quad \Delta \equiv 1 - \delta_C \delta_D q_C q_D.$$

Differentiating with respect to q_C and q_D delivers

$$\frac{\partial \alpha^{CF}}{\partial q_C} = \frac{\delta_D(\delta_C q_D - 1)}{\Delta^2} < 0 \qquad \text{and} \qquad \frac{\partial \alpha^{CF}}{\partial q_D} = \frac{\delta_C \delta_D q_C (1 - \delta_D q_C)}{\Delta^2} > 0,$$

where the inequalities follow from $\delta_C, \delta_D \in (0,1)$ and $q_C, q_D \in (0,1]$.

On the other hand, since $q_C(\theta) = 1 - \lambda_C(\theta)$ and $q_D(\theta) = 1 - \lambda_D(\theta)$, then $q_C'(\theta) = -\lambda_C'(\theta) > 0$ and $q_D'(\theta) = -\lambda_D'(\theta) < 0$ hold. Therefore, by the chain rule

$$\frac{d\alpha^{CF}(\theta)}{d\theta} = \frac{\partial \alpha^{CF}}{\partial q_C} q_C'(\theta) + \frac{\partial \alpha^{CF}}{\partial q_D} q_D'(\theta) < 0,$$

because $\partial \alpha^{CF}/\partial q_C < 0$ and $q_C' > 0$ make the first term negative, while $\partial \alpha^{CF}/\partial q_D > 0$ and $q_D' < 0$ make the second term negative.

By Proposition 1, $\alpha^{DF}(\theta) = \delta_C q_D(\theta) \alpha^{CF}(\theta)$. Differentiating we get

$$\frac{d\alpha^{DF}(\theta)}{d\theta} = \delta_C \Big(q_D'(\theta) \alpha^{CF}(\theta) + q_D(\theta) \alpha^{CF\prime}(\theta) \Big).$$

Since $\delta_C > 0$, $q_D'(\theta) < 0$, $\alpha^{CF}(\theta) > 0$, $q_D(\theta) > 0$, and we have already shown $d\alpha^{CF}(\theta)/d\theta < 0$, both terms in parentheses are negative. Hence $d\alpha^{DF}/d\theta < 0$ for all θ. □

A.3.3 Proof of Proposition 3

Before proving the proposition, we use the following lemma.

Lemma 1 (Monotone entry values)**.** *Fix ω and treat S_1 as given (equivalently, fix F^{CIP}). Then the unmatched value $J_i^U(\theta,\omega)$ is continuous and weakly decreasing in θ for every i.*

Proof. Fix ω and suppress it from notation. Let $F^{\text{CIP}} \equiv \frac{S_1 R^d}{R^f}$ and write $\alpha(\theta) \equiv \alpha^{(p)}(\theta) \in (0,1)$. Define $g_i \equiv \gamma_i \text{var}(S_2) > 0$ and $c(\theta) \equiv \frac{\kappa(\theta,\omega)}{2(R^f)^2} > 0$.

From Eq. (7) in Corollary 1, the within-match notional reads as

$$X_i(\theta) = \frac{\alpha(\theta)}{g_i + \alpha(\theta)c(\theta)}\big(F^{\text{CIP}} - F_i^C\big). \tag{A.11}$$

Total match surplus (evaluated at the within-match equilibrium) is $S_i(\theta) = X_i(\theta)\big(F^D(X_i(\theta);\theta) - F_i^C\big)$. Using $F^D = F^{\text{CIP}} - c(\theta)X_i(\theta)$ and Eq. (A.11), we obtain

$$S_i(\theta) = \frac{\alpha(\theta)\, g_i}{\big(g_i + \alpha(\theta)c(\theta)\big)^2}\big(F^{\text{CIP}} - F_i^C\big)^2. \tag{A.12}$$

By definition, the client-side surplus upon matching is $\Sigma_{C,i}^{(p)}(\theta) = \alpha(\theta)\, S_i(\theta)$ and reads as

$$J_i^U(\theta) = \frac{\delta_C}{1-\delta_C}\, p_C(\theta)\, \frac{\alpha(\theta)^2\, g_i}{\big(g_i + \alpha(\theta)c(\theta)\big)^2}\big(F^{\text{CIP}} - F_i^C\big)^2. \tag{A.13}$$

Continuity follows from continuity of $p_C(\cdot)$, $\alpha(\cdot)$, and $\kappa(\cdot,\omega)$ given by Assumptions 1 and 2 and Proposition 1.

To show monotonicity, note that $p_C(\theta)$ is strictly decreasing in θ. Moreover, the term $\left(\frac{\alpha(\theta)}{g_i+\alpha(\theta)c(\theta)}\right)^2$ is (i) strictly increasing in α and (ii) strictly decreasing in c because, for $g_i > 0$, $\alpha > 0$, and $c > 0$,

$$\frac{\partial}{\partial \alpha}\log\left(\frac{\alpha^2}{(g_i+\alpha c)^2}\right) = \frac{2}{\alpha} - \frac{2c}{g_i+\alpha c} = \frac{2g_i}{\alpha(g_i+\alpha c)} > 0$$
$$\frac{\partial}{\partial c}\log\left(\frac{\alpha^2}{(g_i+\alpha c)^2}\right) = -\frac{2\alpha}{g_i+\alpha c} < 0.$$

By Proposition 2, $\alpha(\theta)$ is strictly decreasing in θ, and by Assumption 1, $c(\theta)$ is strictly increasing in θ. Therefore the product in Eq. (A.13) is weakly decreasing in θ (strictly decreasing whenever $F^{\text{CIP}} \neq F_i^C$), which proves the claim.

□

Now we are ready to prove Proposition 3.

Proof. Fix ω and treat S_1 as given. Define $\bar{\theta} \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i < \infty$. Since $0 \le G(\cdot) \le 1$,

$$0 \le T(\theta;\omega) \le \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i = \bar{\theta} \qquad \text{for all } \theta \ge 0,$$

so $T(\cdot;\omega)$ maps $[0,\bar{\theta}]$ into itself. By Assumption 4, each $\theta \mapsto G(J_i^U(\theta,\omega))$ is continuous, hence $T(\cdot;\omega)$ is continuous on $[0,\bar{\theta}]$.

Define $D(\theta) \equiv T(\theta;\omega) - \theta$. Then D is continuous and satisfies $D(0) = T(0;\omega) \ge 0$ and $D(\bar{\theta}) = T(\bar{\theta};\omega) - \bar{\theta} \le 0$. By the Intermediate Value Theorem, there exists $\theta^* \in [0,\bar{\theta}]$ such that $D(\theta^*) = 0$, i.e. $\theta^* = T(\theta^*;\omega)$.

For uniqueness, let $0 \le \theta' < \theta \le \bar{\theta}$. By Lemma 1, $J_i^U(\theta,\omega)$ is weakly decreasing, and since G is weakly increasing, $G(J_i^U(\theta,\omega))$ is weakly decreasing in θ. Hence $T(\theta;\omega) \le T(\theta';\omega)$ and

$$D(\theta) - D(\theta') = \big(T(\theta;\omega) - T(\theta';\omega)\big) - (\theta - \theta') \le -(\theta - \theta') < 0. \tag{A.14}$$

Thus D is strictly decreasing on $[0,\bar{\theta}]$ and can cross zero at most once. Therefore the fixed point is unique.

□

A.3.4 Proof of Proposition 4

Proof. Fix ω and define $\bar{\theta} \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i < \infty$.

Step 1 (OTC block conditional on S_1). For each $S_1 \in [0,\overline{S}]$, consider the entry fixed point

$$\theta = \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i\, G\big(J_i^U(\theta,\omega;S_1)\big).$$

This is the same object as in Proposition 3, with the only difference that the spot rate S_1 (equivalently $F^{\mathrm{CIP}} = \frac{S_1 R^d}{R^f}$) is treated as a parameter. Under the assumptions stated in Proposition 4 (i), for each $S_1 \in [0,\overline{S}]$ there exists a unique solution $\theta(S_1,\omega) \in [0,\bar{\theta}]$.

Step 2 (continuity of $\theta(S_1,\omega)$). Define the function

$$D(\theta,S_1;\omega) \equiv \frac{1}{V}\sum_{i=1}^{I} \overline{U}_i\, G\big(J_i^U(\theta,\omega;S_1)\big) - \theta.$$

By assumption, $D(\theta,S_1;\omega)$ is continuous on $[0,\bar{\theta}] \times [0,\overline{S}]$.

Let $\{S_{1,n}\}_{n\ge 1}$ be any sequence with $S_{1,n} \to S_1$, and let $\theta_n \equiv \theta(S_{1,n},\omega) \in [0,\bar{\theta}]$ denote the unique solution to $D(\theta_n,S_{1,n};\omega) = 0$. Since $[0,\bar{\theta}]$ is compact, $\{\theta_n\}$ has a convergent subsequence with limit $\widetilde{\theta}$. By continuity of D, $0 = \lim_k D(\theta_{n_k},S_{1,n_k};\omega) = D(\widetilde{\theta},S_1;\omega)$. By uniqueness of the root at S_1, we must have

$\widetilde{\theta} = \theta(S_1, \omega)$. Hence every convergent subsequence of $\{\theta_n\}$ has the same limit, implying $\theta_n \to \theta(S_1, \omega)$. Therefore $S_1 \mapsto \theta(S_1, \omega)$ is continuous.

Step 3 (existence of a joint equilibrium). Define the bond-market excess demand

$$\mathcal{Z}(S_1; \omega) \equiv H_e + H_d(S_1) + H_D(\theta(S_1, \omega), \omega; S_1) - \overline{B}.$$

By continuity of $H_d(\cdot)$ and $H_D(\cdot)$ and by Step 2, $\mathcal{Z}(\cdot; \omega)$ is continuous on $[0, \overline{S}]$. By assumption, there exists $\underline{S} \in [0, \overline{S}]$ such that $\mathcal{Z}(\underline{S}; \omega) \geq 0$. We now show that $\mathcal{Z}(\overline{S}; \omega) \leq 0$.

By definition, $\overline{S} \geq \overline{S}^{\text{CIP}} = \frac{R^f}{R^d}\overline{F}^C$, so $F^{\text{CIP}}(\overline{S}) = \frac{\overline{S}R^d}{R^f} \geq \overline{F}^C \equiv \max_{i \in \mathcal{I}} F_i^C$. By Eq. (7), this implies $X_i(\theta, \omega; \overline{S}) \geq 0$ for all i and all θ, and therefore $\mathcal{X}^{\text{OTC}}(\theta, \omega; \overline{S}) \geq 0$. By the dealer-hedging mapping Eq. (11), we obtain $H_D(\theta, \omega; \overline{S}) \leq 0$.

Moreover, $\overline{S} \geq \overline{S}^{\text{Bd}} = \frac{\text{E}[S_2]R^f - \gamma_d (R^f)^2 \text{var}(S_2)\left(\overline{B} - H_e\right)}{R^d}$ implies $H_e + H_d(\overline{S}) \leq \overline{B}$ by Eq. (12). Combining both inequalities yields $\mathcal{Z}(\overline{S}; \omega) \leq 0$.

Hence, by the Intermediate Value Theorem, there exists $S_1^* \in [\underline{S}, \overline{S}]$ such that $\mathcal{Z}(S_1^*; \omega) = 0$. Let $\theta^* \equiv \theta(S_1^*, \omega)$. By construction, (θ^*, S_1^*) satisfies Eqs. (13) and (14) and is therefore a joint equilibrium.

Step 4 (uniqueness). If $\mathcal{Z}(\cdot; \omega)$ is strictly decreasing on $[\underline{S}, \overline{S}]$, then the root S_1^* is unique. Since for each S_1 the OTC block admits a unique solution $\theta(S_1, \omega)$, the pair (θ^*, S_1^*) is also unique.

To verify strict monotonicity under the sufficient "deep bond market" condition in Proposition 4, note that H_d is affine in S_1 and satisfies $-\frac{d}{dS_1}H_d(S_1) = \frac{R^d}{\gamma_d (R^f)^2 \text{var}(S_2)}$ for all S_1 by Eq. (12). If, in addition, $S_1 \mapsto H_D(\theta(S_1, \omega), \omega; S_1)$ is differentiable and satisfies

$$\sup_{S_1 \in [\underline{S}, \overline{S}]} \frac{d}{dS_1} H_D(\theta(S_1, \omega), \omega; S_1) < \frac{R^d}{\gamma_d (R^f)^2 \text{var}(S_2)},$$

then for all $S_1 \in [\underline{S}, \overline{S}]$, $\frac{d}{dS_1}\mathcal{Z}(S_1; \omega) = \frac{d}{dS_1}H_d(S_1) + \frac{d}{dS_1}H_D(\theta(S_1, \omega), \omega; S_1) < 0$. Therefore $\mathcal{Z}(\cdot; \omega)$ is strictly decreasing and the equilibrium is unique. $\square$