BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 743

An intermediation-based model of exchange rates

by Semyon Malamud and Andreas Schrimpf

Monetary and Economic Department

September 2018

JEL classification: E44, E52, F31, F33, G13, G15, G23

Keywords: Financial intermediation, exchange rates, safe haven, covered interest parity deviations, currency crashes

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

An Intermediation-Based Model of Exchange Rates *

Semyon Malamud† and Andreas Schrimpf‡

This version: September 13, 2018

Abstract

We develop a general equilibrium model with intermediaries at the heart of international financial markets. In our model, intermediaries bargain with their customers and extract rents for providing access to foreign claims. The behavior of intermediaries, by tilting state prices, generates an explicit, non-linear risk structure in exchange rates. We show how this endogenous risk structure helps explain a number of anomalies in foreign exchange and international capital markets, including the safe haven properties of exchange rates and the breakdown of covered interest parity.

Keywords: financial intermediation, exchange rates, safe haven, covered interest parity deviations

JEL Classification Numbers: E44, E52, F31, F33, G13, G15, G23

*We thank Ana Babus, Saki Bigio, Pasquale Della Corte, Michael H. A. Dempster (discussant), Darrell Duffie, Vincent Glode, Tarek Hassan (discussant), Benjamin Hebert, James Kemp, Arvind Krishnamurthy, Felix Kübler, Ye Li, Matteo Maggiori, Frédéric Malherbe, Steve Mobbs, Christian Opp, Monika Piazzesi, Hyun Song Shin, Andrea Vedolin (discussant), Colin Ward (discussant), Pierre-Olivier Weill, Amir Yaron, as well as seminar participants at UCSD, UCLA, Wharton, and conference participants at the 2nd LAEF conference on OTC markets, the Chicago Booth International Macro-Finance Conference, the 2018 Adam Smith Asset Pricing Conference, the 14th Cowles Conference on General Equilibrium, and Stanford Institute for Theoretical Economics for helpful comments. Semyon Malamud acknowledges the financial support of the Swiss National Science Foundation (SNSF Starting Grant "Liquidity") and the Swiss Finance Institute. Parts of this paper were written when Malamud visited the Bank for International Settlements (BIS) as a research fellow. The views in this article are those of the authors and do not necessarily represent those of BIS.

†Swiss Finance Institute, EPF Lausanne, and CEPR; E-mail: semyon.malamud@epfl.ch

‡Bank of International Settlements (BIS) and CEPR; Email: andreas.schrimpf@bis.org

The goal of this paper is to develop a macroeconomic general equilibrium model in which international financial markets are subject to intermediation frictions. In our model, intermediaries use their market power to extract rents from their customers for providing them with access to foreign financial instruments. We show how the behavior of intermediaries, by tilting state prices, generates an explicit, non-linear risk structure in exchange rates and helps account for some of the major anomalies in foreign exchange (FX) and international capital markets, including the safe haven properties of exchange rates and the breakdown of covered interest parity (CIP).

Intermediaries are key to the functioning of international financial markets. The trading of key financial instruments, such as sovereign and corporate bonds, spot FX rates, FX forwards and swaps, and most other derivatives used for hedging purposes, typically occur in over-the-counter (OTC) transactions through financial intermediaries.[1] Trading in such markets is subject to frictions, whereby a handful of global intermediaries exerts significant market power.[2] To study the effect of these market frictions on the macroeconomy and exchange rates, we introduce an imperfectly competitive intermediation sector into a classical cash-in-advance model à-la Lucas (1982). Our model features an economy with multiple countries and partially integrated financial markets. Each country is populated by two classes of agents, customers (households) and specialists (intermediaries). While customers have free access to domestic markets for simple local securities (domestic nominal risk-free bonds and the domestic stock market index), they must use intermediaries to gain access to foreign assets and financial instruments in the dealer-to-customer (D2C) market segment.

[1] Indeed, a large part of the trading in global securities and derivatives markets occurs via over-the-counter transactions, with bank dealers as the major suppliers of intermediation services. For example, daily turnover in interest rate swaps reached almost USD 2 trillion per day in April 2016, while daily trading volume in the global FX market exceeds USD 5 trillion, according to the most recent Bank for International Settlements (BIS) statistics on global OTC derivatives markets. See, BIS (2016). Trading in global OTC markets dwarfs the volume on equities and futures exchanges. In OTC markets, an identical asset is typically traded at different prices at a given point in time, depending on the identity of the trading counterparties.

[2] See, e.g., Hau et al. (2017), who provide evidence for significant rent extraction in the foreign exchange derivatives markets. According to Hau et al. (2017), "A corporate client at the 75th percentile of average transaction costs pays a roughly 12 times larger spread than a corporate client at the 25th percentile."

Intermediaries charge for lending their balance sheet to customers. Upon contact, they take into account customers' optimal demand for foreign financial asset exposures and use their bargaining power to extract rents and charge markups for catering to customers' demand. At the same time, intermediaries use the dealer-to-dealer (D2D) market to manage their balance sheets and share risks arising from trading with customers.

The key implication of our model is an explicit characterization of the (distorted) international risk sharing in terms of a single, country-specific object, $Cov_t(i)$, given by the time-t conditional covariance of country i nominal stock market wealth with country i customers' relative wealth. Here, relative wealth is defined as the ratio of customers' domestic stock market wealth to their (domestic and foreign) total wealth:

$$\text{relative wealth} \equiv \frac{\text{country } i \text{ stock market wealth}}{\text{country } i \text{ customers' wealth}}. \tag{1}$$

We show that the sign of the conditional covariance $Cov_t(i)$ leads to an endogenous (and potentially time-varying) dichotomy of countries in terms of risk sharing and the behavior of exchange rates. The underlying mechanism is entirely determined by the nature of intermediation frictions in our model.[3] Specifically, customers' behavior in the D2C market depends on the diversification benefits gained from the exposure to foreign shocks. In turn, the latter depend on the domestic stock market's ability to span customers' desired wealth profile. This ability is precisely captured by $Cov_t(i)$. We show that when $Cov_t(i)$ is positive (negative), the benefits from international diversification are low (high), and country i customers buy (sell) local nominal bonds from (to) intermediaries. Intermediaries, being short (long) nominal bonds, profit (suffer) from monetary expansions, their net worth goes up (down), and their marginal utilities drop (increase). Since intermediaries are marginal investors in international financial markets, the exchange rate is given by the ratio of intermediary marginal utilities;

[3] Absent frictions, nobody holds nominal assets, and hence monetary policy is neutral and has no impact on international risk sharing.

hence, a drop in the marginal utility of country i intermediaries leads to an excessive exchange rate devaluation. Without sticky prices, the standard export channel is absent and, as a result, exchange rate devaluation always leads to losses in a country's international investment position and a drop in the domestic dollar wealth.

A key implication of our results is that *ex-ante expectations* about future state-contingent monetary policy are crucial for determining the nature of the transmission of monetary shocks and their impact on domestic welfare. The underlying mechanism is as follows: Time-t expectations about future policy determine the nature of (ex-ante) time-t contracts signed in the local D2C market. Intermediaries share the risks from their D2C exposures in the international D2D market. In turn, this leads to an international redistribution of wealth at time $t+1$ when contracts pay off.

One of our goals is to characterize explicitly how monetary policy expectations influence the response of exchange rates to global macroeconomic conditions, in particular, the so-called safe haven properties of exchange rates. The term "safe haven" is commonly used for currencies that tend to appreciate at times when the world marginal utility is high (e.g., US dollar). However, as Maggiori (2013) argues, the large size of the US intermediation sector is difficult to reconcile with the safe haven status of the dollar due to the natural role of the US as global insurance provider. Indeed, this role implies that large wealth transfers from the US to the rest of the world occur during bad times, leading to dollar appreciation during such times. This is what Maggiori (2013) calls the "reserve currency paradox."

We show how intermediation markups may help resolve this paradox. A larger intermediation sector can absorb more risks and, hence, in equilibrium, leads to lower markups because customers have lower overall risk exposure. As a result, intermediaries charge lower markups for insurance against global disaster states and customers respond by acquiring more insurance against these states;[4] ex-post, when a bad state is realized, intermediaries need

[4]In particular, intermediaries end up selling too much insurance, for example, through volatility selling strategies that are popular among real-world intermediaries.

to make larger transfers to customers, their wealth drops, and their marginal utility spikes, making their domestic currency appreciate against currencies with smaller intermediation sectors.

Markups serve as the key barrier to efficient international risk sharing in our model. Understanding the structure of these markups allows us characterize how risk exposures are spread across different countries. We show that markups satisfy a *fundamental markup equation* akin to the fundamental equation of asset pricing. Using this equation, we show that markups have a two-factor structure, with the factors given by the relative wealth (1) and customers' marginal utility of wealth. While the sign of the relative wealth premium is usually positive, the sign of the premium on the marginal utility of wealth coincides with that of $Cov_t(i)$. Thus, while the standard frictionless consumption Euler equation implies that securities that are better able to span the marginal utility of wealth always have higher prices, such securities may or may not have a higher markup, depending on the sign of $Cov_t(i)$.

We illustrate the mechanisms underlying the fundamental markup equation by using it to study one of the most surprising developments in global financial markets over the past few years: the breakdown of covered interest rate parity.[5] In our model, customers willing to borrow or lend in a foreign currency cannot do so directly and must go through intermediaries. For example, they can borrow in the local currency and then enter an FX swap contract with the intermediary to borrow dollars synthetically; the corresponding indirect rate of borrowing dollars may be quite different from the rate at which the intermediaries can borrow dollars directly. This leads to a breakdown of covered interest rate parity between the D2C and D2D market segments and a non-zero *cross-currency dollar basis,* defined as the difference between the two (indirect minus direct) US dollar rates. The above-mentioned

[5] CIP states that the interest rate differentials implicit in foreign exchange swap markets coincide with the corresponding differential in money market rates in the two countries. The breakdown of CIP even for some of the world's most liquid currency pairs is documented in Du et al. (2016), Avdjiev et al. (2016), Borio et al. (2016), and Rime et al. (2017).

two-factor structure of the fundamental markup equation implies that the dollar basis against country i is positive if either (i) the dollar positively co-moves with country i relative wealth 1 or (ii) the dollar co-moves with country i customers' marginal utility of wealth and the sign of that co-movement coincides with that of $Cov_t(i)$. In particular, for countries with a positive $Cov_t(i)$, the safe haven properties of the dollar may lead to a positive basis.

Roadmap. The remainder of the paper is structured as follows. Section 1 provides an overview of the relevant literature. Section 2 describes the model. Section 3 provides the equilibrium characterization. Section 4 investigates the link between intermediation frictions and various exchange rate anomalies. Section 5 concludes.

1 Literature Review

The literature on general equilibrium models of exchange rates is vast. Most papers assume either complete markets (see, e.g., Lucas (1982); Cole and Obstfeld (1991), Dumas (1992); Backus et al. (1992); Backus and Smith (1993); Obstfeld and Rogoff (1995); Pavlova and Rigobon (2007); Verdelhan (2010); Colacito and Croce (2011)) or an exogenously specified incompleteness in the form of portfolio constraints (see, e.g., Chari et al. (2002); Corsetti et al. (2008); Pavlova and Rigobon (2008)), unspanned risk factors (Pavlova and Rigobon (2010, 2012), Farhi and Gabaix (2016), Brunnermeier and Sannikov (2017)), or limits to market participation (Alvarez et al. (2002, 2009), Bacchetta and Van Wincoop (2010) and Hassan (2013)). In contrast, in our model, market incompleteness and limits to international risk sharing are endogenous and determined by equilibrium intermediation markups.

The most closely related to ours are papers by Maggiori (2013), Gabaix and Maggiori (2015).[6] Maggiori (2013) considers a two-country model characterized by asymmetry in financial intermediation capacity: In his model, one country (US) has a better developed

[6]Several papers (see, e.g., Jeanne and Rose (2002), Evans and Lyons (2002), Hau and Rey (2006), Bruno and Shin (2014), Camanho et al. (2017)) study the impact of frictions on exchange rates without modeling

(i.e., less credit constrained) intermediation sector. During global crises, the US suffers heavier losses (through wealth transfers to the rest of the world) because of its role as a global insurer, leading to asymmetric international risk sharing[7] and a "reserve currency paradox." In our paper, we show how intermediation markups can help in resolving this paradox.

Gabaix and Maggiori (2015) develop a general equilibrium model of exchange rates based on the limited risk-bearing capacity of financial intermediaries.[8] In their model, intermediaries demand a risk premium for holding currency risk originating in global imbalances. Gabaix and Maggiori (2015) show that this simple intermediation friction has a major impact on equilibrium FX dynamics; in particular, their model can rationalize many of the important stylized facts about exchange rates and link these stylized facts to intermediaries' balance sheets.[9]

As in Gabaix and Maggiori (2015), in our model, markets have a two-tier structure (D2C and D2D segments), and intermediaries play the key role in absorbing global imbalances; hence, in both Gabaix and Maggiori (2015) and our model, intermediaries' risk-bearing capacity is a key variable driving global risk premiums and exchange rates. However, the nature of frictions in the D2C market is different in our model and arises from intermediary market power and the implied *endogenous* market segmentation. In particular, in contrast

fundamentals such as exports and imports of multiple goods. Instead, they focus on how the behavior and incentive structure of intermediaries shape market outcomes in FX.

[7] Kindleberger (1965), Despres et al. (1966), Caballero et al. (2008), Mendoza et al. (2009), and Chien and Naknoi (2015) also emphasize differences in financial development across countries as an important source of global imbalances.

[8] The importance of intermediation frictions for the transmission and amplification of shocks in domestic markets has been acknowledged in many papers. See, e.g., Holmstrom and Tirole (1997), Bernanke et al. (1999), Gertler and Kiyotaki (2010), He and Krishnamurthy (2011, 2013, 2014), Adrian and Boyarchenko (2012), Brunnermeier and Sannikov (2014, 2016), Adrian et al. (2014), Rampini and Viswanathan (2015), He et al. (2016), Korinek and Simsek (2016), Piazzesi and Schneider (2016), Bianchi and Bigio (2016), Bigio and Sannikov (2016), Malamud and Schrimpf (2016), Malherbe and McMahon (2017), and Coimbra and Rey (2017).

[9] Itskhoki and Mukhin (2017) develop a dynamic model with exogenous small but persistent shocks to international bond markets that could, for example, originate from shocks to intermediary balance sheets. They show that a model with such shocks is quantitatively consistent with the empirically observed joint dynamics of exchange rates and macro variables.

to models with exogenously specified limits to market participation,[10] our model shows how barriers to international trade (intermediation markups) arise endogenously depending on the risk structure of international shocks. Finally, in Gabaix and Maggiori (2015) the dynamics of intermediaries' risk-bearing capacity is specified exogenously, while in our model it is endogenous and fully micro-founded and proportional to intermediaries' net worth. Negative shocks to this net worth occur whenever states against which intermediaries sell a lot of insurance are realized, leading to an international redistribution of wealth.

In our model, intermediation markups charged for insurance against some states of the world can become prohibitively high, making pricing kernels and exchange rates exhibit behavior reminiscent of "rare disasters" (see, e.g., Barro (2006)). As Brunnermeier et al. (2008), Lustig et al. (2011), Jurek and Xu (2014), Farhi et al. (2015), and Farhi and Gabaix (2016) argue, rare disaster risk has a first-order impact on equilibrium exchange rates. Our model implies that exchange rate *disasters may arise endogenously due to intermediation frictions,* but only for countries with a negative $Cov_t(i)$: For such countries, customers make highly risky investments, which may lead to very large wealth losses in a disaster state or currency crash. This opens up an important role for stabilization policies, as in Hassan et al. (2016), whereby expectations about such policies alter $Cov_t(i)$ ex-ante and hence may reduce the probability of disasters ex-post.

Our results about international risk sharing and its link to heterogeneous exposure of different countries to global shocks links our paper to the work of Colacito et al. (2017). Specifically, Colacito et al. (2017) show (both theoretically and empirically) how countries' heterogeneity in exposures to global shocks affects the equilibrium risk structure of exchange rates. We show explicitly how such heterogeneity affects the nature of contracts signed in the D2C market and, as a result, influences international risk sharing.

Finally, our paper is also related to the recent literature on the breakdown of covered interest parity. See, for example, Du et al. (2016), Avdjiev et al. (2016), Borio et al. (2016),

[10] Such as those of Alvarez et al. (2002, 2009), and Bacchetta and Van Wincoop (2010).

and Rime et al. (2017). Several papers derive CIP deviations using models with different forms of limits to arbitrage. For example, Amador et al. (2017) show how CIP deviations arise in a small open economy at the zero lower bound; Ivashina et al. (2015) and Liao (2016) highlight the importance of global banks' demand for dollar funding;[11] Hebert (2017) shows how the cross-section of CIP deviations can be used to recover intermediaries' financial constraints; Andersen et al. (2017) show how seemingly riskless arbitrage (including CIP) may not be economically viable due to dealers' funding value adjustment; and Jiang et al. (2018) link CIP deviations to the special role of US Treasury securities as collateral. To the best of our knowledge, our model is the first multi-country macroeconomic general equilibrium model that generates a breakdown of CIP endogenously through segmentation effects in imperfect international financial markets. In particular, we shed light on the macroeconomic origins of CIP deviations, their signs, and differences across countries.

2 The Model

2.1 Agents, Preferences, and Consumption

We consider an international multiple goods monetary economy with intra-temporal[12] cash-in-advance constraints, as in Lucas (1982). Time is discrete, $t = 0, 1, \cdots, T$, and the information structure is characterized by a probability space (Ω, P) equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$. There are N countries, indexed by $i = 1, \cdots, N$. Country i produces a tradable good, also indexed by i. Country i tradable good is produced by an endowment process

[11] See, also, Aldasoro et al. (2017), who provide evidence that Japanese banks, which are known to have a particularly high demand for US dollar funding, face significant markups when accessing dollar repo markets, consistent with the mechanism highlighted in our model.

[12] That is, agents only need to hold cash within the period for consumption needs and do not store cash inter-temporally. For example, this can be achieved if intermediaries deposit their cash holdings overnight with the central bank, and the bank pays interest on such cash equal to the equilibrium nominal rate. This interest is then simply a part of the total cash rebates to domestic agents.

$X_{i,t},\ i = 1, \cdots, N,\ t \geq 0$. The government of country i controls the supply of domestic, country i currency, $\mathcal{M}_{i,t}$.

Each country is populated by two classes of agents, I-agents (intermediaries or dealers) and H-agents (households or customers) that have identical, time-varying, stochastic time discount factors $\Psi_{i,t}$ (demand shocks), $i = 1, \cdots, N$. Such time-preference shocks are commonly used in international economics. See, for example, Stockman and Tesar (1995), Pavlova and Rigobon (2007), and Itskhoki and Mukhin (2017). We assume that all agents derive utility from consumption $C_{i,t}^{bundle}$ of a country-specific bundle of tradable goods. All agents are endowed with standard, inter-temporal, logarithmic preferences.

$$E\left[\sum_{t=0}^{T} \Psi_{i,t} \log C_{i,t}^{bundle,J}\right],$$

where

$$C_{i,t}^{bundle,J} = \prod_{k=1}^{\mathcal{N}} (C_{i,k,t}^{J})^{\theta_{i,k}},\ i = 1, \cdots, N,\ J = I,\ H$$

is the country-specific tradable goods consumption bundle and $C_{i,k,t}^{I,H},\ k = 1, \cdots, N$ is the time-t consumption of country-k tradable good in country i by the corresponding agents' class I, H. The log preferences assumption implies that we can normalize $\theta_{i,k}$ so that

$$\sum_k \theta_{i,k} = 1,\ i = 1, \cdots, N.$$

We denote by $P_{i,k,t}$ the nominal price of good k in country i, in the units of the local currency. We also denote by $\mathcal{E}_{i,t}$ the US dollar price of the currency of country i; that is, whenever $\mathcal{E}_{i,t}$ goes up, the local currency of country i appreciates against the US dollar. Below, we will always use the currency of country 1 (US dollars) as the reference currency and \$ to denote the corresponding economic variables. Since our focus is on financial market frictions, we abstract from frictions in international goods markets and assume that purchasing power parity always holds.[13] In this case, nominal goods prices in local currencies satisfy

$$\mathcal{E}_{i,t}\, P_{i,k,t} \;=\; \mathcal{E}_{k,t}\, P_{k,k,t}\ ,\ \ i,k = 1,\cdots,N\,.$$

We assume a cash-in-advance constraint à-la Lucas (1982) at the country level: All country k goods need to be purchased with country k currency, implying that total nominal expenditures for country k tradable goods' endowment $X_{k,t}$ always equals country k money supply:

$$P_{k,k,t}\, X_{k,t} \;=\; \mathcal{M}_{k,t}\ ,\ \ k = 1,\cdots,N\ ,\ \ t \geq 0\,. \tag{2}$$

By direct calculation, given the total nominal expenditure in the units of currency i,

$$C_{i,t}^{I,H} \;\equiv\; \sum_{k=1}^{\mathcal{N}} P_{i,k,t} C_{i,k,t}^{I,H}\,,$$

the optimal consumption bundle of the respective agent class is given by

$$C_{i,k,t}^{I,H} \;=\; C_{i,t}^{I,H} P_{i,k,t}^{-1} \theta_{i,k}\,,\ \ i,k = 1,\cdots,N\,.$$

2.2 Financial Market Structure

We assume that all class I agents (intermediaries) from each country $i = 1,\cdots,N$ have direct access to a frictionless, complete, centralized, international D2D market. We interpret these agents as specialists who possess a technology that allows them to issue and trade general state-contingent claims (a full set of Arrow securities) with other I-agents. Since markets are complete, the prices of all financial securities traded in the inter-dealer market

[13] Note that consumption bundles differ across countries and, hence, the law of one price holds at the goods level but not at the price index level.

can be encoded in a single, international US dollar nominal pricing kernel $M^I_{\$,t,t+1}$ quoted in units of currency 1, so that the time-t US dollar price q_t of a state-contingent claim with a dollar payoff Y_{t+1} is given by

$$q_t \ = \ E_t[M^I_{\$,t,t+1} Y_{t+1}] \, .$$

We will refer to $M^I_{\$,t,t+1}$ as the (US dollar) D2D pricing kernel. We will also use $M^I_{i,t,t+1}$ to denote the D2D kernel denominated in country i currency. By no arbitrage and D2D market completeness, we always have

$$M^I_{i,t,t+1} \ = \ M^I_{\$,t,t+1} \mathcal{E}_{i,t+1} / \mathcal{E}_{i,t} \, . \tag{3}$$

In stark contrast to class-I agents, class H agents (henceforth, customers) of a given country i do not have direct access to the inter-dealer market. We assume, however, that each country i has a domestic centralized exchange at which all local customers and all intermediaries can trade two securities: the claim on country i endowment $X_{i,t}$ (the stock index of country i) and a one-period country-i nominal risk-free bond. Customers willing to trade any other financial instrument must contact an intermediary (an intermediation firm) and bargain over the counter in a D2C market. In particular, this means that customers can borrow or lend in their local currency at prevailing market rates, but those that wish to borrow or lend in a foreign currency must do so through intermediaries, domestic or foreign: Since all intermediaries use a common D2D discount factor, markups do not depend on the intermediary's country of origin and, hence, it does not matter which intermediaries customers contact. For simplicity, we assume that information is symmetric and, hence, intermediaries can observe customers' country of origin. Importantly, customers from different countries cannot *directly* trade financial securities with each other.[14]

[14] In particular, trading foreign stocks can also be done only through intermediaries. This assumption allows us to capture the fact that trading and owning foreign stocks often involves significant amounts

Following He and Krishnamurthy (2013), we assume that class-I agents are specialists who run intermediation firms. The objective of such a firm is to maximize the firm value (i.e., the present discounted value of intermediation markups) under the D2D pricing kernel. Since markets are complete, the risk-neutral firm's objective coincides with that of the risk-averse specialists who run it: Indeed, both the firm's and the specialists' objective is to maximize the present value of revenues under the unique D2D pricing kernel. Therefore, in the following, we identify class-I agents with the intermediation firm they run and call them intermediaries.[15] We formalize the details of the bargaining protocol in the following assumption (see Figure 1 for a graphical description of our model's market structure).

Assumption 1 *In the beginning of each period t, each customer of country i is matched with one intermediary and requests quotes for prices of all one-period-ahead state-contingent claims.*[16] *The intermediary quotes a one-period-ahead country-specific D2C pricing kernel $M^H_{i,t,t+1}$ in the local currency and has full bargaining power in choosing $M^H_{i,t,t+1}$ due to search frictions: If the customer rejects the offer, he can trade country i endowment claims and country i one-period risk-free bonds in the country i centralized exchange with other country i customers and (a continuum of) intermediaries but then has to wait one more period to be matched with another intermediary. The quotes are binding: After receiving the quote, the customer chooses an optimal bundle of state-contingent claims and the intermediary sells this bundle to the customer at the quoted price.*

of intermediation. Similarly, short selling a stock (both local and foreign) always involves intermediation, whereby the short seller must go to an intermediary who then needs to locate a stock owner to borrow the stock. See, e.g., Duffie et al. (2005).

[15]For simplicity, we assume that specialists are the only shareholders of intermediaries and, hence, markups are not rebated back to customers: By assumption, customers (class-H agents) can only freely trade claims on their wealth and short-term bonds. This assumption is made for simplicity and can be relaxed. Allowing customers to freely trade intermediary stocks would add another Lagrange multiplier to the shadow costs of intermediation and thus complicate the analysis.

[16]The assumption of trading only one-period claims with intermediaries is standard in the literature. As Brunnermeier and Koby (2016) argue, this is without loss of generality if old contracts are indexed on contemporaneous economic conditions.

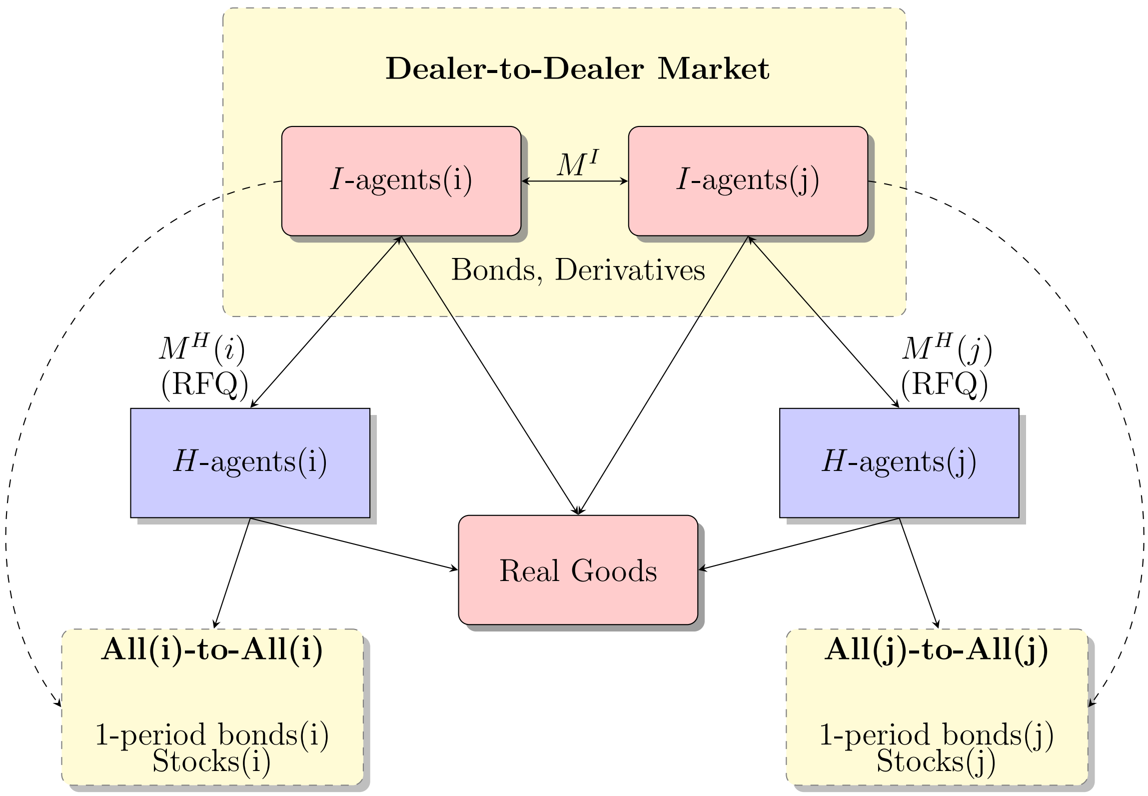

Figure 1: Graphical description of market structure in our model for the two-country case (country i and country j). RFQ denotes the request-for-quote protocol commonly used in D2C segments of OTC markets.

The key mechanisms in our model depend crucially on intermediaries' ability to extract rents. The assumption of monopolistic competition is made for tractability reasons and can be relaxed; for example, our results can easily be adjusted to allow for a different bargaining protocol with a bargaining power below one, such as the Nash protocol that is commonly used in the literature on OTC markets. See, Duffie et al. (2005), Duffie et al. (2007)) Lagos and Rocheteau (2009), and Atkeson et al. (2015).[17] However, some papers (see, e.g., Petersen and Rajan (1995)) argue that monopolistic competition in the intermediation sector is a closer approximation to reality due to switching and relationship costs. See, also, Sharpe (1997), Kim et al. (2003), Bolton et al. (2016), Brunnermeier and Koby (2016), Duffie and Krishnamurthy (2016), and Acharya and Plantin (2016).

2.3 D2C Bargaining and Markups

Assumption 1 implies that we can formulate the bargaining problem in terms of the local currency nominal D2C state prices $M^H_{i,t,t+1}$, quoted by the country i intermediary to a country i customer.[18] We will use $r_{i,t}$ to denote the short-term country i nominal interest rate and let $S_{i,t}$ denote the nominal D2D market price of the total endowment, $\{X_{i,\tau}\}_{\tau=t}^{T}$, of the local, country-$i$ good. By (2), the nominal value of time$-\tau$ endowment equals the time$-\tau$ money supply for any $\tau \geq 0$, and, hence, $S_{i,t}$ coincides with the present value of the stream of future country i money supply, $\mathcal{M}_{i,\tau},\ \tau \geq t$. Hereafter, we interpret this claim as the country-i

[17] The new regulatory environment (based on the Dodd-Frank Act) is designed to move bilateral relationship trading to electronic platforms. For example, trading of standardized interest rate swaps in the US has to a large extent moved to so-called swap execution facilities (SEFs). An all-to-all market such as in equities markets remains a distant reality, though. Most D2C transactions are executed via an RFQ protocol, which is equivalent to an electronic form of OTC trading. The original two-tier market structure thus shows remarkable persistence, with a D2D segment at the core of the market, as in our model. The same is true for fixed income and FX markets. See Collin-Dufresne et al. (2016), Bech et al. (2016), and Moore et al. (2016).

[18] Hebert (2017) investigates a model with a form of market segmentation that is similar to that assumed in our paper. Specifically, Hebert (2017) considers an incomplete market model in which intermediaries can trade a full set of state-contingent claims with each other in the D2D market, while households are constrained in the set of assets they can trade with each other and with intermediaries, who are facing convex portfolio constraints. As a result of this segmentation, Hebert's model also features two pricing kernels, as well as deviations from the law of one price.

stock index and call it the local stock price. By assumption, local customers can freely trade the endowment claim as well as one-period nominal risk-free bonds with other customers and intermediaries in the centralized local exchange for domestic nominal bonds and the domestic stock index. Since competitive intermediaries can freely trade in both the centralized local exchange and the global D2D market, in equilibrium, they will equalize bond and stock prices across these two markets. Hence, domestic nominal bonds and the domestic stock index will trade at D2D prices on the domestic exchange. Thus, customers can trade these two securities at D2D prices: The nominal bond price, $E_t[M^I_{i,t,t+1}]$ (recall (3)) and the nominal stock price $S_{i,t} = \mathcal{M}_{i,t} \ + \ E_t[M^I_{i,t,t+1} S_{i,t+1}]$. Now, since customers can conduct the arbitrage trade in between the centralized domestic exchange and the binding quotes they receive in the D2C market, the intermediary has to quote fair prices for both instruments. Otherwise, customers would immediately arbitrage away the differences, leading to unbounded losses for the intermediary. Formally, this means that the D2C nominal pricing kernel $M^H_{i,t,t+1}$ quoted by the intermediary must satisfy two constraints (fair pricing of domestic bonds and fair pricing of domestic stocks) relating $M^H_{i,t,t+1}$ to the D2D nominal pricing kernel (3) in the local currency:

$$E_t[M^H_{i,t,t+1}] \ = \ E_t[M^I_{i,t,t+1}] \tag{4}$$

$$E_t[M^H_{i,t,t+1} S_{i,t+1}] \ = \ E_t[M^I_{i,t,t+1} S_{i,t+1}] \,. \tag{5}$$

We will also make the following assumption.

Assumption 2 *We assume that class I and class H agents in country i are endowed with the respective shares α_i and $1-\alpha_i$ of the total endowment of the country i tradable good. At time zero, intermediaries pay a cost $\bar{K}_{i,0}$ to customers to set up intermediation firms. The monetary authority controls the money supply through direct rebates to intermediaries.*[19] *We*

[19]For example, as Brunnermeier and Sannikov (2016) argue, controlling the rate on the central bank reserves is effectively equivalent to controlling the supply of central bank money, whereby interest payments on

also use $\mathcal{N}_{i,t+1} \equiv \mathcal{M}_{i,t+1}/\mathcal{M}_{i,t}$ *to denote the growth in money supply for all* $i = 1, \cdots, N\,,\ t \geq 0\,.$

We will use $W^J_{i,t},\ J = I, H$ to denote the wealth of class-J agents. By Assumption 2, customers' time zero nominal net worth is given by $W^H_{i,0} = (1-\alpha_i)S_{i,0} + \bar{K}_{i,0}$.[20] Since markets are complete, customers can use trading in the D2C market to attain any state-contingent consumption expenditures profile $(C^H_{i,t})_{t\geq 0}$ *in the local currency* satisfying the inter-temporal budget constraint:

$$W^H_{i,t} \ = \ C^H_{i,t} \ + \ E_t[M^H_{i,t,t+1} W^H_{i,t+1}]\,,\ W^H_{i,t} \ \geq \ 0$$

for all t. That is, wealth is split between current consumption and the value of financial investments, which in turn equals the present value of future wealth. Denote by

$$D_{i,t} \ \equiv \ E_t\left[\sum_{\tau=0}^{T-t} \frac{\Psi_{i,t+\tau}}{\Psi_{i,t}}\right]\,,$$

the expected discount factor for the whole future consumption stream. Standard results imply that the following is true.

Lemma 1 *Country* i *customers' and intermediaries' nominal consumption and nominal wealth dynamics are given by*

$$\frac{C^{I,H}_{i,t}}{C^{I,H}_{i,t-1}} \ = \ (M^{I,H}_{i,t-1,t})^{-1}\,\frac{\Psi_{i,t}}{\Psi_{i,t-1}} \tag{6}$$

reserves are equivalent to direct money rebates to intermediaries. Note, however, that market segmentation implies that the distribution of money holdings has real effects in our model and, hence, cannot be neglected.

[20]By assumption, $\bar{K}_{i,0}$ is immediately transferred to customers at time zero. The assumption that the cost is only incurred at time zero is made for convenience and can be relaxed. These costs will play no role in the subsequent analysis. One could potentially use them to endogenize the size of the intermediation sector as well as to study the impact of regulations on the endogenous size of the intermediation sector and markups. Importantly, making these costs sufficiently large, we can make intermediary net worth, $W^I_{i,0}$, arbitrarily small. They also allow us to make an important distinction between the size of markups and the actual profitability of the intermediation sector: While the markups (i.e., the spread between the D2C and the D2D pricing kernels) might be high, the actual profit margins might be quite low.

and

$$\frac{W_{i,t}^{I,H}}{W_{i,t-1}^{I,H}} = (M_{i,t-1,t}^{I,H})^{-1} \frac{\Psi_{i,t}}{\Psi_{i,t-1}} \frac{D_{i,t}}{D_{i,t-1}}, \; i = 1, \cdots, N.$$

If $D_{i,t}$ is stochastic, then our model features a non-constant consumption/wealth ratio.

The intuition behind Lemma 1 is straightforward: A log utility maximizing agent always consumes inversely proportionally to state prices (Equation (6)). Furthermore, his decision to allocate wealth across states is driven by the product of the discount factor $\Psi_{i,t+1}$ and the expected discount factor D_{t+1}; the latter determines the value of the total future stream of consumption in a given state.

Let us now consider the bargaining problem between a customer and an intermediary. At time t, a country i customer with nominal wealth $W_{i,t}^H$ gets matched with an intermediary who quotes him a one-period-ahead pricing kernel $M_{i,t,t+1}^H$ in the local currency. Given this quote, the customer decides how to optimally finance his future wealth $W_{i,t+1}^H$ through a portfolio of the risk-free bond and the stock to be traded in the domestic centralized exchange, as well as an OTC contract acquired in the D2C market, with a potentially complex state-contingent payoff. Due to no-arbitrage constraints (4)-(5) for the local stock and bond markets, customers are in fact indifferent between trading the stock and bond at the domestic exchange or in the D2C market. Hence, without loss of generality, we can assume that they directly trade bonds and stocks with intermediaries. Thus, a customer is simply buying the claim on his future wealth, $W_{i,t+1}^H$.

Formula (6) implies that intermediaries are facing a downward-sloping demand curve from customers, state by state: $W_{i,t+1}^H(M_{i,t,t+1}^H) = W_{i,t}^H \frac{\Psi_{i,t+1} D_{i,t+1}}{\Psi_{i,t} D_{i,t}} (M_{i,t,t+1}^H)^{-1}$. Since intermediaries have access to complete D2D markets, their objective is to maximize the present value of cash flows in the D2C market under the D2D pricing kernel. Those cash flows are given by $E_t[M_{i,t,t+1}^H W_{i,t+1}^H]$ (the price paid by customers to intermediaries) at time t and by

$-W^H_{i,t+1}$ (the contractual payments of intermediaries to customers) at time $t+1$. Thus, their present value under the D2D pricing kernel is given by the total markup

$$E_t[M^H_{i,t,t+1} W^H_{i,t+1}(M^H_{i,t,t+1})] - E_t[M^I_{i,t,t+1} W^H_{i,t+1}(M^H_{i,t,t+1})] \,. \tag{7}$$

This is the difference between the values of the claim $W^H_{i,t+1}$ under the D2C and the D2D pricing kernels. The intermediary's goal is thus to maximize (7) under the no-arbitrage constraints (4)-(5). Denoting by $\mu_{i,t}$ and $\lambda_{i,t}$ the Lagrange multipliers for constraints (4) and (5), respectively, and writing down the first-order conditions for the intermediary's maximization problem *state-by-state*, we get

$$M^I_{i,t,t+1} \frac{\Psi_{i,t+1}\, D_{i,t+1}}{\Psi_{i,t} D_{i,t}} (M^H_{i,t,t+1})^{-2} \;=\; \lambda_{i,t}(S_{i,t+1}/S_{i,t}) + \mu_{i,t} \,. \tag{8}$$

The intuition behind (8) is as follows: The marginal gain of selling insurance against a state x is given by the product of the D2D price $M^I_{i,t,t+1}(x)$ and the sensitivity of the customer's consumption to the price $M^H_{i,t,t+1}(x)$. Since customers have log utility, this sensitivity is given by $-\Psi_{i,t,t+1}\, D_{i,t,t+1}\, (M^H_{i,t,t+1})^{-2}$. At the optimum, this marginal gain is equal to the *state-contingent* shadow cost of constraints (4)-(5), given by $\lambda_{i,t}(S_{i,t+1}/S_{i,t}) + \mu_{i,t}$.

Denote

$$\Psi_{i,t,t+1} \;\equiv\; \frac{\Psi_{i,t+1}}{\Psi_{i,t}},\;\; D_{i,t,t+1} \;\equiv\; \frac{D_{i,t+1}}{D_{i,t}},\;\; S_{i,t,t+1} \;\equiv\; S_{i,t+1}/S_{i,t}\,,\;\; W^H_{i,t,t+1} \;\equiv\; \frac{W^H_{i,t+1}}{W^H_{i,t}}\,.$$

The solution to (8) is given by the following proposition.

Proposition 2 *The optimal pricing kernel quoted by the intermediary is given by*

$$M^H_{i,t,t+1} \;=\; \frac{(\Psi_{i,t,t+1}\, D_{i,t,t+1})^{1/2} (M^I_{i,t,t+1})^{1/2}}{(\lambda_{i,t} S_{i,t,t+1} + \mu_{i,t})^{1/2}}\,, \tag{9}$$

where the Lagrange multipliers $\lambda_{i,t}$, $\mu_{i,t} \in \mathbb{R}$ *are given by*

$$\begin{aligned} \lambda_{i,t} &= \frac{\mathrm{Cov}_t^H(-(W_{i,t,t+1}^H)^{-1}, S_{i,t,t+1})}{\Delta_{i,t}} \\ \mu_{i,t} &= \frac{\mathrm{Cov}_t^H((S_{i,t,t+1}/W_{i,t,t+1}^H),\, S_{i,t,t+1})}{\Delta_{i,t}} \end{aligned} \tag{10}$$

with

$$\Delta_{i,t} \;=\; E_t^H[(W_{i,t,t+1}^H)^{-1} S_{i,t,t+1}^2]\, E_t^H[(W_{i,t,t+1}^H)^{-1}] - \left(E_t^H[(W_{i,t,t+1}^H)^{-1} S_{i,t,t+1}]\right)^2 \;>\; 0\,,$$

and where $E_t^H[\cdot]$ *and* $\mathrm{Cov}_t^H(\cdot)$ *denote, respectively, the conditional expectation and covariance under the conditional D2C risk-neutral measure,* $E_t[M_{i,t,t+1}^H]^{-1} M_{i,t,t+1}^H$.

Proposition 2 is key to the subsequent analysis. It shows how the bargaining friction and intermediaries' ability to charge state-contingent markups affects asset prices and, as a result, distorts equilibrium allocations. Indeed, since customers and intermediaries have identical log preferences, their consumption profiles should be proportional to each other; however, distorted state prices (9) imply that customers' consumption will be inefficiently low (high) relative to that of intermediaries when D2C state prices are high (low) relative to D2D state prices.

The signs of $\lambda_{i,t},\ \mu_{i,t}$ will play an important role in the subsequent analysis. Since these quantities are Lagrange multipliers of constraints (4)-(5), their signs are determined by the "direction" in which these constraints are binding. Consider the Lagrange multiplier $\mu_{i,t}$ of the constraint (4). One can equivalently interpret (4) as a pair of inequality constraints

$$\begin{aligned} &E_t[M_{i,t,t+1}^H] \ge (1-\varepsilon) E_t[M_{i,t,t+1}^I] \\ &E_t[M_{i,t,t+1}^H] \le (1+\varepsilon) E_t[M_{i,t,t+1}^I]\,, \end{aligned}$$

where the parameter ε (determining the corridor inside which the intermediary can quote rates) is arbitrarily small.

The economic intuition behind these constraints is as follows. If customers want to invest in risk-free assets[21], the intermediary will try to push the nominal rate $e^{r_{i,t}} = 1/E_t[M^H_{i,t,t+1}]$ all the way down to its lower bound, determined by the D2D market rate $1/E_t[M^I_{i,t,t+1}]$, and, hence, the constraint $E_t[M^H_{i,t,t+1}] \leq E_t[M^I_{i,t,t+1}]$ will be binding; in this case, standard Kuhn-Tucker conditions imply that $\mu_{i,t} > 0$. This is what we call a *"risk-off" scenario* or a "flight-to-safety" episode.

In contrast, if customers find it optimal to borrow from intermediaries, the latter will try to push the offered rate all the way up to its upper bound, determined by the D2D market rate $1/E_t[M^I_{i,t,t+1}]$, and, hence, the constraint $E_t[M^H_{i,t,t+1}] \geq E_t[M^I_{i,t,t+1}]$ will be binding; in this case, standard Kuhn-Tucker conditions imply that $\mu_{i,t} < 0$. This is what we call a *"risk-on" scenario* or a "search-for-yield" episode.

Explicit expressions (10) show that the signs and the magnitudes of the multipliers are determined by the stock market's ability to hedge fluctuations in customers' wealth. Under normal circumstances, wealth and stock prices are positively related and, hence, $\lambda_{i,t} > 0$. A positive $\lambda_{i,t}$ leads to an endogenous home bias: Since customers' wealth growth, $W_{i,t,t+1}$, is inversely proportional to $M^H_{i,t,t+1}$, it loads positively on $S_{i,t+1}$. Thus, in our model, markups lead to an excessive exposure to the local stock market. At the same time, the sign of $\mu_{i,t}$ coincides with that of

$$\mathrm{Cov}_t(i) \;\equiv\; \mathrm{Cov}^H_t((S_{i,t,t+1}/W^H_{i,t,t+1}),\, S_{i,t,t+1})\,. \tag{11}$$

When $\mathrm{Cov}_t(i)$ is negative, stock market returns are high in the states in which stocks under-

[21] Intermediaries can provide access to (nearly) risk-free assets through private money creation, for example, through bank deposits and money market funds. We abstract from such private money creation in our model. See, Brunnermeier and Sannikov (2016) for a model featuring an impact of such private money creation on monetary policy passthrough.

perform relative to other asset classes composing customers' wealth. For example, stock returns are procyclical, but customers' wealth has a high embedded leverage, implying that this wealth exhibits larger swings with the business cycle. Such a risk profile precisely means that stocks are not sufficient to satisfy customers' search for yield. Customers respond to this risk profile by borrowing from intermediaries and taking leveraged bets on foreign shocks.[22] The opposite happens when $\mathrm{Cov}_t(i)$ is positive. In particular, time variations in $\mathrm{Cov}_t(i)$ may lead to an endogenous customer leverage cycle (see, e.g., Geanakoplos (2010)).

2.4 The Fundamental Markup Equation

We assume that domestic stocks and domestic nominal bonds are not sufficient to span the *non-linear* consumption profile desired by customers.[23] As an illustration, consider a eurozone customer who wants to buy insurance against a major depreciation of the euro (EUR). While this might theoretically be done by buying a deep out-of-the-money (OTM) put option on the euro, in reality such contracts are not traded on organized exchanges, neither centralized nor inter-dealer platforms, and even receiving a quote from a major dealer may be difficult, in particular, because it is difficult to determine the fair value of this contract. Such a deal would have to be settled through OTC bargaining and would involve significant markups. While the probability of a major EUR depreciation is low, the marginal utility of European customers in this disaster state is high; thus, markups for such OTM options may have large impacts on asset prices.

Consistent with this intuition, the key role of intermediaries in our model is to offer customers access to securities with non-linear payoffs because linear securities spanned by domestic stocks and bonds cannot do so. Customers' demand for non-linear asset payoffs

[22] For example, it is well known (see, e.g., Maggiori (2013)) that the United States benefits from large yields on its foreign assets relative to foreign liabilities.

[23] For this to be true, we need enough (at least three) states at each node of the event tree. In particular, the country-specific degree to which local stocks and bonds are able to span customers' target risk profile determines the country-specific magnitude of the impact of intermediation markups on risk sharing.

determines the size and the sign of the security's markups, defined as the spread between the price in the D2C and D2D market segments: Formally, for an asset with a nominal payoff Y_{t+1} in the local currency, the markup paid by country i customers is given by

$$E_t[M^H_{i,t,t+1}Y_{t+1}] - E_t[M^I_{i,t,t+1}Y_{t+1}]\,. \tag{12}$$

Using (4), we can rewrite (12) as follows:

Proposition 3 *[The fundamental markup equation] Intermediation markups paid by country−i customers on a nominal claim Y_{t+1} denominated in country-i currency are given by*

$$Markup_{i,t}(Y_{t+1}) \;=\; -e^{-r_{i,t}}\,\mathrm{Cov}^H_t\left(\frac{M^I_{i,t,t+1}}{M^H_{i,t,t+1}}\,,\,Y_{t+1}\right)\,. \tag{13}$$

Equation (13) is the *fundamental markup equation*, akin to the fundamental equation of asset pricing that characterizes risk premiums through the covariance with the stochastic discount factor. It originates from the *risk-based* approach to markups in our model, whereby the risk properties of an asset's payoff are key determinants of the magnitude of markups, in both the time series (due to the dynamics of the conditional covariance (13)) and the cross-section. It is these markups that distort equilibrium allocations and drive all of our main results.

Log utility and Lemma 1 together with (8) imply that the pricing kernel quotient equals

$$\frac{M^I_{i,t,t+1}}{M^H_{i,t,t+1}} \;=\; \lambda_{i,t}(S_{i,t,t+1}/W^H_{i,t,t+1}) + \mu_{i,t}(W^H_{i,t,t+1})^{-1}$$

Hence, using (10) and (11), we can decompose markups (13) as

$$\begin{aligned} Markup_{i,t}(Y_{t+1}) \;=\; &-\; e^{-r_{i,t}}\,\lambda_{i,t}\mathrm{Cov}^H_t\left((S_{i,t,t+1}/W^H_{i,t,t+1})\,,\,Y_{t+1}\right) \\ &-\; e^{-r_{i,t}}\,\frac{Cov_t(i)}{\Delta_{i,t}}\mathrm{Cov}^H_t\left((W^H_{i,t,t+1})^{-1}\,,\,Y_{t+1}\right)\,. \end{aligned} \tag{14}$$

Thus, markups have a two-factor structure, and a security's markup is determined by two components: its ability to hedge shocks to the ratio $S_{i,t,t+1}/W^H_{i,t,t+1}$ and to the marginal utility of wealth, $(W^H_{i,t,t+1})^{-1}$. By (10), under normal circumstances, we have $\lambda_{i,t} > 0$, and, hence, a security always gets a higher markup when Y_{t+1} can be used as a hedge against states in which customers' wealth drops relative to the stock market. In contrast, the security's ability to serve as a hedge against states with low wealth may have an ambiguous impact on the markup, depending on the sign of $\frac{\mu_{i,t}}{\lambda_{i,t}} = \frac{\mathrm{Cov}_t(i)}{\mathrm{Cov}^H_t(-(W^H_{i,t,t+1})^{-1}, S_{i,t,t+1})}$. When $\frac{\mu_{i,t}}{\lambda_{i,t}} < 0$ (search-for-yield episode), securities offering insurance against low-wealth states become expensive; this provokes customers to underinsure against these states and to make excessive bets on the upside. In contrast, when $\frac{\mu_{i,t}}{\lambda_{i,t}} > 0$ (flight-to-safety episode), insurance against low-wealth states becomes cheap, and customers overinsure against these states.

3 Equilibrium

Equilibrium prices are pinned down by imposing market clearing for all goods. By the cash-in-advance constraint (2), total nominal expenditures for country k goods equal the money supply and, hence, using (6), we get

$$\sum_{i=1}^{N} (C^H_{i,t} + C^I_{i,t})\, \mathcal{E}_{i,t}\, \theta_{i,k} \;=\; \mathcal{M}_{k,t}\mathcal{E}_{k,t}\ ,\ \ k = 1, \cdots, N\,.$$

In general, the structure of equilibrium can be quite complex and depends in a non-trivial way on the distribution of preference parameters $\theta_{i,k}$ across countries. To isolate the local demand/supply effects from those of global demand and supply, we will assume that consumption demand exhibits a single-factor structure, so that

$$\theta_{i,k} \;=\; \bar{\theta}_k \beta_i + (1 - \beta_i)\, \delta_{i,k}\,,$$

where $\delta_{i,k}$ is the Kronecker delta and $\sum_{k=1}^{N} \bar{\theta}_k = 1$. Here, $1 - \beta_i$ measures the degree of consumption home bias in country i, while $\bar{\theta}_k$ reflects the global demand for country k goods. Thus, β_i measures how much country i participates in global trade. In the sequel, we will refer to $(1-\beta_i)^{-1}\beta_i$ as the trade weight of country i. Let $\bar{B} \equiv \sum_k (1-\beta_k)^{-1}\beta_k\theta_k$, and define

$$Dollar_t \equiv -(1+\bar{B})^{-1} \sum_j (1-\beta_j)^{-1}\beta_j \mathcal{M}_{j,t}\mathcal{E}_{j,t}\,, \tag{15}$$

to be the global, trade-weighted dollar index. When the US dollar appreciates relative to other currencies, $\mathcal{E}_{j,t}$ drop, and the dollar index $Dollar_t$ rises in value. By direct calculation, the simple unit-elastic preference specification implies that global consumption risk sharing is linear, and a country k nominal consumption expenditures (in dollars), $(C^H_{k,t} + C^I_{k,t})\mathcal{E}_{k,t}$, are given by a linear combination of domestic output value, $\mathcal{M}_{k,t}\mathcal{E}_{k,t}$, and $Dollar_t$, as characterized in the following lemma.

Lemma 4 *Total dollar consumption expenditures of country k agents is given by*

$$(C^H_{k,t} + C^I_{k,t})\mathcal{E}_{k,t} = (1-\beta_k)^{-1}\left(\mathcal{M}_{k,t}\mathcal{E}_{k,t} + \theta_k Dollar_t\right). \tag{16}$$

Having computed country k consumption, we can now proceed to determining pricing kernels and exchange rates. Specifically, substituting (9) into Lemma 1, we get a quadratic equation for the D2D pricing kernel, which can be solved explicitly.

Proposition 5 *Let*

$$Y_{\$,t+1} \equiv 0.5(1-\beta_\$)C^H_{\$,t}(\lambda_{\$,t}S_{\$,t,t+1} + \mu_{\$,t})^{1/2}(D_{\$,t,t+1})^{-1/2}\,.$$

Then,

$$M^I_{\$,t,t+1}(Dollar_{t+1}) \;=\; \Psi_{\$,t,t+1}\left(\frac{Y_{\$,t+1}+\sqrt{Y^2_{\$,t+1}+(\mathcal{M}_{\$,t+1}\;+\;\theta_\$ Dollar_{t+1})(1-\beta_\$)C^I_{\$,t}}}{(\mathcal{M}_{\$,t+1}\;+\;\theta_\$ Dollar_{t+1})}\right)^2. \tag{17}$$

Formula (17) shows explicitly how intermediation frictions enter the equilibrium pricing kernel and distort the nature of risk premiums. In the frictionless model, consumption is perfectly aligned across customers and intermediaries and, hence, (16) implies that the frictionless pricing kernel, $M^{I,*}_{\$,t,t+1}$, is given by

$$M^{I,*}_{\$,t,t+1} \;=\; \Psi_{\$,t,t+1}(C^H_{\$,t}\;+\;C^I_{\$,t})(\mathcal{M}_{\$,t+1}\;+\;\theta_\$ Dollar_{t+1})^{-1}.$$

It only loads on three risk factors: the time discount factor $\Psi_{\$,t,t+1}$, money supply $\mathcal{M}_{\$,t+1}$, and $Dollar_{t+1}$.[24] This is intuitive: An increase in the US money supply reduces the value of the dollar, while an increase in the dollar factor signifies a drop in the global consumption demand. Formula (17) shows that intermediation frictions introduce a non-linear, multi-factor structure into the pricing kernel. In addition to the simple effects present in the frictionless case, the pricing kernel loads on the domestic stock market, $S_{\$,t+1}$, and the wealth-consumption ratio, $D_{\$,t+1}$. The loadings on these two additional factors depend on the size and the signs of the shadow costs $\lambda_{i,t},\ \mu_{i,t}$.

Substituting the country-specific pricing kernels into the equation

$$\mathcal{E}_{k,t,t+1} \;=\; \frac{M^I_{k,t,t+1}}{M^I_{\$,t,t+1}},$$

we get a non-linear fixed-point equation for $\mathcal{E}_{k,t+1}$. As we show in the Appendix, this equation

[24] Note that the dollar plays no special role in our model: The EUR pricing kernel depends on the trade-weighted euro index.

always has a unique solution. Then, substituting this solution into the defining equation (15) for *Dollar*, we can complete the determination of equilibrium exchange rates as a function of stock returns, $(S_{i,t,t+1})_{i=1}^{N}$. We formalize these results in the following theorem.

Theorem 6 (Non-linear Risk Structure of Exchange Rates) *Define*

$$Y_{k,t+1} \;\equiv\; 0.5(1-\beta_k)C^H_{k,t}(\lambda_{k,t}S_{k,t,t+1}+\mu_{k,t})^{1/2}(D_{k,t,t+1})^{-1/2}\,,$$

and let $g_{k,t+1}(z)$ *be the unique positive solution to the implicit equation*

$$g_{k,t+1}(z) \;=\; \Psi_{k,t,t+1}\left(\frac{Y_{k,t+1}+\sqrt{Y^2_{k,t+1}+(\mathcal{M}_{k,t+1}\;+\;\theta_k\mathcal{E}^{-1}_{k,t}z/g_{k,t+1}(z))(1-\beta_k)C^I_{k,t}}}{(\mathcal{M}_{k,t+1}\;+\;\theta_k\mathcal{E}^{-1}_{k,t}z/g_{k,t+1}(z))}\right)^2.$$

Then,

$$\mathcal{E}_{k,t,t+1} \;=\; \frac{g_{k,t+1}(Dollar_{t+1}M^I_{\$,t,t+1}(Dollar_{t+1}))}{M^I_{\$,t,t+1}(Dollar_{t+1})}\,, \tag{18}$$

while the dollar index, $Dollar_{t+1}$*, is a solution to the fixed-point equation*

$$\begin{aligned}&M^I_{\$,t,t+1}(Dollar_{t+1})Dollar_{t+1}\\ &=\; -(1+\bar{B})^{-1}\sum_{k=1}^{N}(1-\beta_k)^{-1}\beta_k\mathcal{M}_{k,t+1}\mathcal{E}_{k,t}g_{k,t+1}(Dollar_{t+1}M^I_{\$,t,t+1}(Dollar_{t+1}))\,.\end{aligned}$$

Theorem 6 shows how intermediation frictions give rise to an endogenous, non-linear risk structure in exchange rates. Since intermediaries are marginal investors in international markets, exchange rates' changes are given by the ratios of their marginal utilities:

$$\mathcal{E}_{k,t,t+1} \;=\; \frac{\Psi_{k,t,t+1}}{\Psi_{\$,t,t+1}}\frac{(C^I_{k,t,t+1})^{-1}}{(C^I_{\$,t,t+1})^{-1}}\,,$$

By (16), intermediaries' consumption is determined by two forces: aggregate country k

consumption, as given by $\mathcal{M}_{k,t}\mathcal{E}_{k,t} + \theta_k Dollar_t$, and the nature of sharing this consumption risk between customers and intermediaries. In turn, the nature of this customer-intermediary risk sharing is determined by intermediation markups.

Specifically, when a state with a low shadow cost of intermediation, $\lambda_{k,t}S_{k,t,t+1}+\mu_{k,t}$, is realized, customers make large transfers to intermediaries, customers' consumption drops, and intermediary net worth goes up, leading to a currency depreciation. Such opposite behavior of customers' and intermediaries' marginal utilities is consistent with the Backus and Smith (1993) puzzle; in fact, consumption marginal utilities are often negatively correlated with exchange rates (see, e.g., Backus and Kehoe (1992)), in stark contrast with what frictionless models predict. Similarly, when a state with a high wealth-consumption ratio is realized, customers' consumption is low relative to their future wealth and currency depreciates.

The link between exchange rates and the wealth-consumption ratio is similar to that in models with recursive preferences (see, e.g., Colacito and Croce (2011)); however, in our model it arises endogenously due to intermediation markups. Importantly, and in stark contrast to most existing asset pricing models, both the wealth-consumption ratio, $D_{k,t+1}$, and the stock market return, $S_{k,t,t+1}$, appear in the pricing kernel and the exchange rates, and the pricing kernel loads positively on the latter and negatively on the former. In particular, intermediation markups lead to endogenous and potentially time-varying correlations between exchange rates and the stock market.

Similar to models with recursive preferences, formula (17) (as well as its counterpart (18) for exchange rates) is merely an equilibrium relationship between the pricing kernel and the market return. Thus, solving for an equilibrium boils down to finding the return on the domestic stock market, $S_{k,t,t+1}$, the dollar index $Dollar_{t+1}$, and the shadow costs $\lambda_{i,t},\ \mu_{i,t}$. Theorem 6 shows how $Dollar_{t+1}$ can be derived explicitly in terms of the world stock market returns $(S_{k,t,t+1})_{k=1}^N$, and the latter can be used to find $\lambda_{i,t},\ \mu_{i,t}$ from formulas (10). Thus, finding an equilibrium reduces to finding $(S_{k,t,t+1})_{k=1}^N$. While this cannot generally be done in closed form, a lot can be learned about the structure of $(S_{k,t,t+1})_{k=1}^N$ from just analyzing implications of international goods' market clearing.

To this end, we will denote by $S^{\$}_{k,t}$ the US dollar price of country k stock market: $S^{\$}_{k,t} = S_{k,t}\mathcal{E}_{k,t}$. We will also define

$$\bar{S}^{\$}_t \;\equiv\; (1+\bar{B})^{-1}\sum_i (1-\beta_i)^{-1}\beta_i S^{\$}_{i,t}$$

to be the US dollar price of the *global, trade-weighted stock market portfolio.* This portfolio will play a key role in our analysis. It is this, the trade-weighted portfolio, and not the international (market capitalization-weighted) market portfolio that will determine the factor structure of international stock returns. Multiplying the consumption market clearing condition (16) by the stochastic discount factor $M^I_{k,\tau,\tau+1}$, summing over τ, taking expectation, and using the fact that an agent's wealth equals the present value of his consumption, we arrive at the following result:

Proposition 7 *We have*

$$S^{\$}_{k,t} \;=\; \theta_k \bar{S}^{\$}_t \;+\; (1-\beta_k)(W^H_{k,t} \;-\; PV_t(Markups_k) \;+\; W^I_{k,t})\mathcal{E}_{k,t}\,, \tag{19}$$

where $PV_t(Markups_k)$ is the present D2D market value of markups (12) paid by customers to intermediaries,

$$PV_t(Markups_k) \;=\; E_t\left[\sum_{\tau=t}^{T-1} M^I_{k,t,\tau} E_t[(M^H_{k,\tau,\tau+1} - M^I_{k,\tau,\tau+1})W^H_{k,\tau+1}]\right].$$

The intuition behind (19) is as follows: $S^{\$}_{k,t}$ is the dollar market value of future output of country k. By market clearing, it also equals the market value of global expenditures for country k goods. The latter are given by the sum of domestic and foreign consumption. By formula (16), foreign agents spend $-\theta_k Dollar_t$ dollars on country k goods, and the present value of these expenditures is precisely given by $\theta_k \bar{S}_t^{\$}$. Domestic consumption can in turn be split into intermediary and customer consumption. The present value of the former is intermediary wealth. However, the present value of customer consumption under the D2D pricing kernel is strictly smaller than customer wealth, and the gap is precisely given by the present value of markups.

Substituting (19) into (10) and (11), we can decompose $Cov_t(k)$ as follows:

$$\begin{aligned}\mathrm{Cov}_t(k) \;=\; & \frac{W_{k,t}^{H,\$}}{S_{k,t}}\Bigg(\theta_k \mathrm{Cov}_t^H\left(\frac{\bar{S}_{t+1}^{\$}}{W_{k,t+1}^{H,\$}},\, S_{k,t,t+1}\right) \\ & + \;\mathrm{Cov}_t^H\left(\frac{W_{k,t+1}^I - PV_{t+1}(Markups_k)}{W_{k,t+1}^H},\, S_{k,t,t+1}\right)\Bigg),\end{aligned} \tag{20}$$

where $W_{k,t}^{H,\$}$ is customers' dollar wealth. As we explain above (see Section 2.4), the sign of $\mathrm{Cov}_t(k)$ determines customers' incentives to take on exposure to risks in international financial markets. Decomposition (20) allows us to identify the general equilibrium forces driving these incentives. The first component is the most interesting one. It shows how customers' risk taking depends on the correlation structure in the international stock market. When $\mathrm{Cov}_t^H\left(\frac{\bar{S}_{t+1}^{\$}}{W_{k,t+1}^{H,\$}},\, S_{k,t,t+1}\right)$ is negative, international stock markets offer attractive diversification benefits for country k customers; when this effect is strong enough, $\mathrm{Cov}_t(k)$ becomes negative, and customers borrow from intermediaries to get international exposure. In contrast, when $\mathrm{Cov}_t^H\left(\frac{\bar{S}_{t+1}^{\$}}{W_{k,t+1}^{H,\$}},\, S_{k,t,t+1}\right)$ is positive, there are no such benefits, and customers instead sell foreign exposure to intermediaries and invest the proceeds in risk-free bonds. The intuition behind the second component, $\mathrm{Cov}_t^H\left(\frac{W_{k,t+1}^I - PV_{t+1}(Markups_k)}{W_{k,t+1}^H},\, S_{k,t,t+1}\right)$, is similar. That is, when $\mathrm{Cov}_t^H\left(\frac{W_{k,t+1}^I - PV_{t+1}(Markups_k)}{W_{k,t+1}^H},\, S_{k,t,t+1}\right)$ is negative, customers' wealth under-reacts to business cycle shocks, when compared to intermediary wealth. Customers respond by contacting intermediaries to get levered exposure to business cycle shocks, and intermediaries charge markups for lending their balance sheets to customers.

4 Applications

To proceed further and derive explicit analytical results, we will follow the approach of Itskhoki and Mukhin (2017) and study the behavior of prices and exchange rates in the limit of a substantial home bias, corresponding to the case when β_i is small for all i.[25]

In our model, country i customers are only willing to get exposure to foreign shocks if there is the possibility for international trade in real goods. Indeed, if $\beta_i = 0$ for all i, then customers simply consume their endowment and there are no gains from trading, in either goods or financial claims.[26] Thus, intermediaries have nothing to charge markups for, and intermediation frictions do not have any impact on equilibrium prices. We will use the * superscript to denote equilibrium objects in the $\beta_i \to 0$ limit. In this limit, exchange rates and stock prices are proportional to money supply:[27]

$$\frac{\mathcal{E}^*_{j,t+1}}{\mathcal{E}^*_{j,t}} \;\equiv\; \frac{\mathcal{N}^{-1}_{j,t+1}\Psi_{i,t,t+1}}{\mathcal{N}^{-1}_{\$,t+1}\Psi_{\$,t,t+1}}\;,\;\; S^*_{i,t} \;=\; \mathcal{M}_{i,t}\,D_{i,t}\,. \tag{21}$$

A substantial home bias economy is a small perturbation of this autarky economy. Trade in goods and securities is approximately linear in β_i, and, hence, so are all equilibrium quantities.

Lemma 8 *In the substantial home bias approximation, we have*

$$\bar{S}^{\$,*}_t \;\approx\; \sum_j \beta_j\, S^{\$,*}_{j,t}$$

[25] As Itskhoki and Mukhin (2017) argue, many countries exhibit significant home bias in consumption. We have also solved the opposite limit of vanishing home bias and most of our results qualitatively hold in this environment. We therefore expect that our results are robust to the degree of the home bias.

[26] Recall that we assume that customers and intermediaries have identical discount factors and, therefore, there are no gains from trading between these two groups within one country.

[27] See Internet Appendix for a formal derivation.

whereas $\lambda_{i,t} \approx 1$, *while*

$$Cov_t(i) \;=\; \hat{\theta}_i \mathrm{Cov}_t^H\left(\frac{\bar{S}^{\$,*}_{t+1}}{S^{\$,*}_{i,t+1}}, S^*_{i,t+1}\right). \tag{22}$$

with some $\hat{\theta}_i > 0$.

The most important result in this lemma is formula (22): As we show in the Appendix, in the substantial home bias approximation, the second component in the decomposition (20) is proportional to the first; thus, customers' risk taking is determined by only one factor: the covariance of $S^*_{i,t+1}$ with the quotient $\frac{\bar{S}^{\$}_{t+1}}{S^{\$}_{i,t+1}}$. That is, it is determined by the quotient of the global, trade-weighted stock market wealth and the domestic stock market wealth. This is intuitive: With substantial home bias, domestic wealth is dominated by domestic stock market wealth and, hence, customers' incentives to get exposure to foreign shocks are determined by the diversification benefits offered by the foreign stock market.

4.1 Stabilization policies and a dichotomy of countries

To obtain further analytical results, we will assume that all variables in our model are driven by a single common shock. Specifically, we will make the following assumption.

Assumption 3 *There exists a Markov process* $\omega_t \in \mathbb{R}$, $t \geq 0$ *with the mean* $\bar{\omega}$ *and a small variance such that* $\Psi_{i,t,t+1}$ *and* $D_{i,t+1}$ *are smooth and monotonic functions* ω_{t+1}.[28] *Furthermore, the government adjusts the money supply in response to shocks:*

$$\log \mathcal{N}_{i,t+1} \;=\; -\alpha_i^{\mathcal{N}}\, \delta_i \omega_{t+1} \;+\; \varepsilon^{\mathcal{N}}_{i,t+1}\,,$$

[28]See the Internet Appendix.

where we have defined

$$\delta_i \;\equiv\; \frac{\partial}{\partial \omega_t} \log D_{i,t}|_{\omega_t=\bar{\omega}} \;>\; 0\,,$$

and where $\alpha_i^{\mathcal{N}}$ *is the sensitivity of the country-*i *monetary policy to global shocks, while* $\varepsilon_{i,t+1}^{\mathcal{N}}$ *stand for idiosyncratic monetary shocks that are independent across* t *and* i*. Also let*

$$\delta_i^{\mathcal{N}} \;\equiv\; \frac{\partial}{\partial \omega_{t+1}} S_{i,t,t+1}^{*}|_{\omega_{t+1}=\bar{\omega}} \;=\; \delta_i(1-\alpha_i^{\mathcal{N}})\,. \tag{23}$$

We say that country i *pursues a stabilization policy if* $\alpha_i^{\mathcal{N}} > 0$*; this stabilization policy is mild (respectively, strong) if* $\alpha_i^{\mathcal{N}} < 1$ *(respectively,* $\alpha_i^{\mathcal{N}} > 1$*).*[29] *We also say that country* j *has less policy uncertainty than country* i *if* $\mathrm{Var}_t[e^{\varepsilon_{j,t+1}^{\mathcal{N}}}] < \mathrm{Var}_t[e^{\varepsilon_{i,t+1}^{\mathcal{N}}}]$.

We will also employ the following definition.

Definition 9 *A country* i *has a high (respectively, low) sensitivity to global shocks if*

$$\delta_i^{S} \;\equiv\; \frac{\partial}{\partial \omega_{t+1}} \frac{S_{i,t,t+1}^{\$}}{\bar{S}_{t,t+1}^{\$}}|_{\omega_{t+1}=\bar{\omega}} \tag{24}$$

is positive (respectively, negative).

Based on the definition of the *global trade-weighted stock market portfolio*, we have

$$\sum_j \frac{\beta_j}{1-\beta_j}\frac{S_{j,t}^{\$}}{\bar{S}_t^{\$}} \;=\; 1 \;\Rightarrow\; \sum_j \frac{\beta_j}{1-\beta_j}\delta_j^{S} \;=\; 0\,.$$

Thus, the average (trade-weighted) sensitivity of countries' relative stock prices (in US

[29]Note that, importantly, the fact that $\mathcal{N}_{i,t+1}$ is monotone decreasing in shocks does not mean that the monetary policy directly reacts to the stock market (even though this does seem to be the case; see, Cieslak and Vissing-Jorgensen (2017)); rather, the policy simply reacts to common macroeconomic shocks. See also Law et al. (2017) for evidence that US monetary policy does react strongly to the macroeconomic situation. We follow Hassan et al. (2016) in using the term "stabilization policy" to describe such a state-contingent monetary policy.

dollars), $\frac{S_{j,t}^{\$}}{S_t^{\$}}$, is zero and, hence, we can classify countries according to their exposure to global shocks, relative to that of the *global trade-weighted stock market portfolio*. By (23) and (24), when the variance of ω_t is sufficiently small, we have

$$\text{sign}(Cov_t(i)) \;=\; -\,\text{sign}(\delta_i^S)\;\text{sign}(1-\alpha_i^{\mathcal{N}})\,.$$

Thus, customers in countries that have high or low sensitivity to global shocks exhibit opposite behavior in their demand for insurance against these shocks; similarly, aggressive stabilization policies will also have opposite effects in such countries. As we show in the Internet Appendix, the total country i US dollar wealth, $(W_{i,t+1}^H + W_{i,t+1}^I)\mathcal{E}_{i,t+1}$, decreases in country i monetary shocks, $\mathcal{N}_{i,t+1}$,[30] if and only if $Cov_t(i) > 0$. The reason is that the dichotomy of Definition 9 directly transmits into an analogous dichotomy for the insurance demand in the D2C market and the opposing reactions of countries to monetary policy. In other words, ex-ante expectations about the state-contingent monetary policy, as captured by $\alpha_i^{\mathcal{N}}$, affect customers' ex-ante insurance demand; this leads to an ex-post wealth distribution across countries when intermediaries make transfers in the global D2D market.

4.2 Safe haven currencies

A large body of literature has investigated differences in the stochastic properties of exchange rates across countries and, in particular, the tendency of some currencies to appreciate in bad times. Several explanations for this behavior have been proposed, including differences in intermediaries' risk-bearing capacity (Gourinchas et al. (2010) and Maggiori (2013)), country size (Martin (2012), Hassan (2013)), factor endowments (Ready et al. (2017), Powers (2015)),

[30] We interpret a positive shock to money supply, $\mathcal{N}_{i,t+1} > 1$, as monetary policy easing: Since the frictionless exchange rates scale inversely with the money supply, this is consistent with empirical evidence (see Ferrari et al. (2017)) indicating that monetary policy easing shocks indeed usually leads to an immediate contemporaneous depreciation of exchange rates.

sensitivity to disaster risk (Farhi and Gabaix (2016)), trade centrality (Richmond (2015)), and exposure to long-run risk (Colacito et al. (2017)). We will employ the following definition:

Definition 10 *We say that currency i is a safe haven relative to currency j if, conditional on time$-t$ information and absent monetary shocks, relative exchange rate changes, $\mathcal{E}_{i,t,t+1}/\mathcal{E}_{j,t,t+1}$ co-move negatively with the global stock market.*

Everywhere in the sequel, we will refer to

$$w_i^* \equiv W_{i,0}^I / W_{i,0}^H$$

as the country i intermediation capacity. This quantity captures the net worth of intermediaries relative to customers and, hence, measures intermediaries' ability to take on balance sheet risk when trading with customers. For simplicity, throughout this subsection, we assume that countries differ only in w_i^* and in their degree of monetary policy uncertainty (see Assumption 3). The following is true.

Proposition 11 *Suppose that $\Psi_{i,t} = \Psi_{\$,t}$ and that countries $i, \$$ have low sensitivity to global shocks. Then, the US dollar is a safe haven currency relative to country i if*

- *Countries only differ in the degrees of policy uncertainty and there is less uncertainty in US policy.*
- *Countries only differ in intermediation capacities and the US has a larger capacity.*

In the frictionless economy, monetary policy uncertainty is irrelevant in the setup of Proposition 11: Absent differences in time discount factors $\Psi_i, \Psi_\$$, exchange rates (21) satisfy $\frac{\mathcal{E}_{i,t+1}}{\mathcal{E}_{i,t}} = \frac{\mathcal{N}_{\$,t+1}}{\mathcal{N}_{i,t+1}}$. That is, when the US money supply grows (relative to that of country i), the US dollar depreciates (relative to country i currency). Hence, exchange rate movements are driven purely by monetary shocks and do not correlate at all with fundamental shocks ω.

Intermediation frictions break this neutrality result, as Proposition 11 shows. Once again, the underlying mechanism relies on the expectations channel. Customers anticipating higher policy uncertainty contact intermediaries to buy insurance against future shocks. Intermediaries charge markups for providing this insurance, in turn limiting customers' ability to efficiently allocate consumption across future states and to buy insurance against global stock market crashes. Put differently, customers in *countries with greater monetary policy uncertainty are less able to insure against global shocks, and their consumption is more sensitive to these shocks.* When a global shock hits, intermediaries in all countries suffer and see their balance sheets shrink, while marginal utilities go up. However, intermediaries in countries with greater policy uncertainty suffer *less* that those in the US because they have sold less insurance against those global crisis states. This means that the exchange rate – given by the ratio of intermediaries' marginal utilities – depreciates relative to the safer US dollar.[31]

As in Gourinchas et al. (2010) and Maggiori (2013), the US dollar is special in our model because US intermediaries act as insurance providers to the rest of the world. However, this means that the US dollar should also suffer the most during crisis periods, leading to what Maggiori (2013) called the "reserve currency paradox." Proposition 11 shows how intermediation markups may help in resolving this puzzle. The underlying mechanism is as follows. Customers in country i suffer in the states with a high quotient $\frac{\bar{S}^{\$}_{t+1}}{S^{\$}_{i,t+1}}$ of the global trade-weighted wealth to domestic wealth because their net worth drops relative to the rest of the world. This is exploited by intermediaries who charge high prices for insurance against these states, pushing state prices up; the size of this effect decreases in their intermediation capacity. As a result, state prices in the country with a larger intermediation sector are less sensitive to the quotient $\frac{\bar{S}^{\$}_{t+1}}{S^{\$}_{i,t+1}}$, making their exchange rates appreciate when this quotient drops. Note finally that the result of Proposition 11 depends on the assumption that the US

[31]This result is consistent with the empirical findings in Mueller et al. (2017), who show that the US dollar tends to depreciate when US monetary policy uncertainty is high.

has low sensitivity to global shocks. The latter is equivalent to a US stock market having lower than average exposure to the world (trade-weighted) stock market portfolio.

4.3 Violations of covered interest parity

In the real world, customers in a foreign country often do not have access to direct dollar borrowing and lending. This also often applies to major foreign regional banks, which are considered to be lower tier in international capital markets and treated as customers by the main global dealer banks. Instead, they can obtain US dollars through intermediaries by borrowing funds in the local currency and then swapping their position into dollars. In a fictitious perfect market, the forces of arbitrage imply that the corresponding FX swap-implied dollar rate should be equal to the dollar rate when directly borrowing funds in US money markets. This arbitrage relationship is known as the covered interest parity condition. However, growing empirical literature (see, e.g., Du et al. (2016), Avdjiev et al. (2016), Borio et al. (2016), and Rime et al. (2017)) provides strong evidence for large and persistent CIP deviations across a multitude of currencies.

In our model, market segmentation naturally leads to a violation of the CIP relationship because customers willing to enter the FX swap position need to do this through intermediaries in an OTC market with non-competitive prices. The FX swap rate quoted by the intermediary will contain a markup.[32] Our goal here is to study the potential macroeconomic drivers of such markups. Recall that $\mathcal{E}_{i,t}$ is the dollar price of country i currency and, hence, $\mathcal{E}_{i,t}^{-1}$ is the currency$-i$ price of the dollar. Denote by $r_{\$,t}^{I,H}$ the "synthetic" nominal US dollar interest rate quoted by intermediaries to customers of country i. By definition,

$$e^{-r_{\$,t}^{I,H}} \;=\; E_t\left[M_{i,t,t+1}^{H}\,\mathcal{E}_{i,t,t+1}^{-1}\right].$$

[32]The global FX swap markets are highly concentrated, with trading dominated by a handful of dealers. See, e.g., Moore et al. (2016).

We define the cross-currency basis against the US dollar as

$$Basis^{\$}_{i,t} \;=\; e^{-r_{\$,t}} - e^{-r^{I,H}_{\$,t}} \;\approx\; r^{I,H}_{\$,t} - r_{\$,t}\,.$$

That is, the basis (as defined here) is given by the difference between the synthetic US dollar rate $r^{I,H}_{\$,t}$ that country i customers can back out from intermediaries' quotes in the FX swap market and the direct dollar rate $r_{\$,t}$, is generally non-zero. Since the CIP relation holds separately in the D2C and D2D markets, country i customers receive from intermediaries a quote $r^{I,H}_{\$,t} \;=\; -\log E_t\left[M^H_{i,t,t+1}\mathcal{E}^{-1}_{i,t,t+1}\right]$, for the dollar rate, as well as a quote $f^{I,H}_{\$,t} \;=\; \log\frac{E_t\left[M^H_{i,t,t+1}\mathcal{E}^{-1}_{i,t+1}\right]}{E_t\left[M^H_{i,t,t+1}\right]}$ for the dollar forward rate. Both are in the D2C market and satisfy the no-arbitrage CIP relation: $r^{I,H}_{\$,t} = \log \$_{i,t} - f^{I,H}_{\$,t} + r_{i,t}$. At the same time, the same relation holds in the D2D market: $r_{\$,t} = \log \$_{i,t} - f^{I,i}_{\$,t} + r_{i,t}$. Thus, the basis is given approximately by

$$Basis^{\$}_{i,t} \;\approx\; r^{I,H}_{\$,t} - r_{\$,t} \;=\; (\log \$_{i,t} - f^{I,H}_{\$,t} + r_{i,t}) \;-\; (\log \$_{i,t} - f^{I,i}_{\$,t} + r_{i,t}) \;=\; f^{I,i}_{\$,t} \;-\; f^{I,H}_{\$,t}\,.$$

Hence, it arises exclusively from the difference in the forward rates, $f^{I,i}_{\$,t} - f^{I,H}_{\$,t}$, and we can interpret $r^{I,H}_{\$,t}$ as the forward-implied rate.

The above mentioned empirical papers show that the basis is positive for a large set of currencies. This means that *borrowing in USD is more expensive in the swap market that in direct USD money markets*, which are not accessible to many customers around the world. Our goal here is to understand the macroeconomic origins of the positive basis. By the fundamental markup equation (14), we have

$$\begin{aligned} Basis^{\$}_{i,t} \;=\;& e^{-r_{i,t}}\,\lambda_{i,t}\mathrm{Cov}^H_t\left((S_{i,t,t+1}/W^H_{i,t,t+1})\,,\,\mathcal{E}^{-1}_{i,t,t+1}\right) \\ &+\; e^{-r_{i,t}}\,\frac{Cov_t(i)}{\Delta_{i,t}}\,\mathrm{Cov}^H_t\left((W^H_{i,t,t+1})^{-1}\,,\,\mathcal{E}^{-1}_{i,t,t+1}\right)\,, \end{aligned} \tag{25}$$

and, hence, a positive basis emerges in our model if the dollar is a good hedge for either the wealth ratio, $(S_{i,t,t+1}/W^H_{i,t,t+1})$, or the marginal utility of wealth, $(W^H_{i,t,t+1})^{-1}$, multiplied by $Cov_t(i)$. As we explain above, these quantities are determined by the sensitivities of countries $i, \$$ to global shocks as well as by the degrees of monetary stabilization policies pursued by the governments, $\alpha_i^{\mathcal{N}},\ \alpha_{\$}^{\mathcal{N}}$. The following proposition derives sufficient conditions for the positivity of the basis in the substantial home bias limit.

Proposition 12 *Suppose that the variances of all shocks are sufficiently small. If* $\delta_k^S(\delta_k^S - \delta_k^{\mathcal{N}}) > 0,\ k = i, \$$, *then a positive basis emerges if countries differ in only one of the following:*

(1) Country i has smaller intermediation capacity than the US.

(2) The US has a higher market capitalization than country i.

Furthermore, the basis is monotone increasing $\alpha_{\$}^{\mathcal{N}}$ *if and only if* $\delta_{\$}^S < 0$.

The key implication of the fundamental markup equation (25) is that the basis is determined by risk properties of the dollar. In case (1) of Proposition 12, the US dollar is attractive because high intermediation capacity makes exchange rates less sensitive to shocks (see the Internet Appendix). A good example for case (2) is an emerging market economy with a stock market that is highly sensitive to global demand shocks (e.g., because its exports of commodities are not sufficiently diversified). Customers in such a country depend crucially on the global demand for their goods; when the US market is a key component of this demand, customers find it optimal to borrow in US dollars. Finally, our last result implies that a positive basis may arise when the US monetary policy starts diverging from that of the rest of the world. Empirical evidence (see Wu and Zhang (2018)) seems to suggest that this is indeed the case.

5 Conclusions

We introduce an imperfectly competitive intermediation sector into a standard, international monetary model à-la Lucas (1982). We show that one simple friction, whereby intermediaries exploit their market power and charge endogenous markups for providing customers with access to foreign securities, can generate the rich behavior of risk premiums and exchange rates. We derive the risk structure of exchange rates explicitly and link it to observable macro-economic quantities.

We show how our explicitly derived risk structure of exchange rates can help explain several known anomalies in exchange rates, such as safe haven properties of exchange rates and the breakdown of covered interest parity. Customers' demand for exposure to foreign shocks is driven by a new object: the covariance of the domestic nominal stock market wealth and the ratio of this wealth to the global trade-weighted stock market wealth. Expectations about the future state-contingent conduct of monetary policy influences this covariance and, hence, affects the nature of international risk sharing.

Our model is flexible and can easily be modified to account for more complex features of the real world. In particular, incorporating sticky prices and realistic monetary policy and inflation dynamics is crucial for understanding the impact of intermediation frictions on macroeconomic dynamics. In a production economy with financially constrained firms, intermediation frictions would also play the additional role of directly affecting the real side of the economy. Furthermore, intermediaries themselves are subject to numerous regulatory requirements and financial frictions. We leave all these important questions for future research.

References

Acharya, Viral and Guillaume Plantin, "Monetary Easing and Financial Instability. Working paper.," 2016.

Adrian, Tobias and Nina Boyarchenko, "Intermediary leverage cycles and financial stability. Working paper.," 2012.

___ , Erkko Etula, and Tyler Muir, "Financial Intermediaries and the Cross-Section of Asset Returns," *Journal of Finance*, 2014, (6), 2557–2596.

Aldasoro, Inaki, Thorsten Ehlers, and Egemen Eren, "The Japan Repo Premium and US Dollar Intermediation," 2017. Working paper.

Alvarez, Fernando, Andrew Atkeson, and Patrick J Kehoe, "Money, Interest Rates, And Exchange Rates With Endogenously Segmented Markets," *Journal of political Economy*, 2002, *110* (1), 73–112.

___ , ___ , and ___ , "Time-varying Risk, Interest Rates, And Exchange Rates In General Equilibrium," *The Review of Economic Studies*, 2009, *76* (3), 851–878.

Amador, Manuel, Javier Bianchi, Luigi Bocola, and Fabrizio Perri, "Exchange rate policies at the zero lower bound," Technical Report, NBER 2017.

Andersen, Leif, Darrell Duffie, and Yang Song, "Funding Value Adjustments," *Journal of Finance*, 2017.

Atkeson, Andrew G, Andrea L Eisfeldt, and Pierre-Olivier Weill, "Entry and exit in OTC derivatives markets," *Econometrica*, 2015, *83*, 2231–2292.

Avdjiev, Stefan, Wenxin Du, Catherine Koch, and Hyun Song Shin, "The dollar, bank leverage and the deviation from covered interest parity," Working Paper, BIS 2016.

Bacchetta, Philippe and Eric Van Wincoop, "Infrequent Portfolio Decisions: A Solution To The Forward Discount Puzzle," *The American Economic Review*, 2010, *100*, 870–904.

Backus, David K and Gregor W Smith, "Consumption And Real Exchange Rates In Dynamic Economies With Non-Traded Goods," *Journal of International Economics*, 1993, *35*, 297–316.

Backus, David K. and Patrick J. Kehoe, "International Evidence on the Historical Properties of Business Cycles," *American Economic Review*, 1992, *82*, 864–888.

Backus, David K, Patrick J Kehoe, and Finn E Kydland, "International Real Business Cycles," *Journal of political Economy*, 1992, *100*, 745–775.

Barro, Robert, "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 2006, *121*, 823–866.

Bech, Morten L, Annamaria Illes, Ulf Lewrick, and Andreas Schrimpf, "Hanging up the phone-electronic trading in fixed income markets and its implications," *BIS Quarterly Review, March 2016*, 2016.

Bernanke, Ben S, Mark Gertler, and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," *Handbook of macroeconomics*, 1999, *1*, 1341–1393.

Bianchi, Javier and Saki Bigio, "Banks, Liquidity Management and Monetary Policy," Technical Report, UCLA 2016.

Bigio, Saki and Yuliy Sannikov, "Credit, Money, Interest, and Prices.," Working Paper, UCLA 2016.

BIS, *Triennial Central Bank Survey. Foreign Exchange Turnover in April 2016*, Basel: Bank for International Settlements, 2016.

Bolton, Patrick, Xavier Freixas, Leonardo Gambacorta, and Paolo Emilio Mistrulli, "Relationship and transaction lending in a crisis," *Review of Financial Studies*, 2016, p. hhw041.

Borio, Claudio, Robert McCauley, Patrick McGuire, and Vladyslav Sushko, "The failure of covered interest parity: FX hedging demand and costly balance sheets," Working paper, BIS 2016.

Brunnermeier, Markus K and Yann Koby, "The Reversal Interest Rate: The Effective Lower Bound on Monetary Policy. Working paper.," 2016.

___ **and Yuliy Sannikov**, "A Macroeconomic Model With A Financial Sector," *The American Economic Review*, 2014, *104* (2), 379–421.

___ **and** ___ , "The I theory of money. Working paper.," 2016.

Brunnermeier, Markus K. and Yuliy Sannikov, "International Monetary Theory: Mundell-Fleming Redux," Technical Report, Princeton University 2017.

Brunnermeier, Markus K, Stefan Nagel, and Lasse H Pedersen, "Carry Trades And Currency Crashes," *NBER macroeconomics annual*, 2008, *23* (1), 313–348.

Bruno, Valentina and Hyun Song Shin, "Cross-Border Banking And Global Liquidity," *The Review of Economic Studies*, 2014. (epub ahead of print), doi: 10.1093/restud/rdu042.

Caballero, Ricardo J, Emmanuel Farhi, and Pierre Olivier Gourinchas, "An equilibrium model of global imbalances and low interest rates," *American Economic Review*, 2008, *98* (1), 358–393.

Camanho, Nelson, Harald Hau, and Helene Rey, "Global Portfolio Rebalancing Under the Microscope," 2017. Working paper.

Chari, Varadarajan V, Patrick J Kehoe, and Ellen R McGrattan, "Can Sticky Price Models Generate Volatile And Persistent Real Exchange Rates?," *The Review of Economic Studies*, 2002, *69*, 533–563.

Chien, Yi-Li and K. Naknoi, "The risk premium and long-run global imbalances," *Journal of Monetary Economics*, 2015, *76*, 299–315.

Cieslak, Anna and Annette Vissing-Jorgensen, "The economics of the "Fed put"," Technical Report, Duke University 2017.

Coimbra, Nuno and Hélène Rey, "Financial Cycles with Heterogeneous Intermediaries," Working Paper 2017.

Colacito, Riccardo and Mariano M Croce, "Risks For The Long Run and The Real Exchange Rate," *Journal Of Political Economy*, 2011, *119*, 153–181.

___ **, Mariano Massimiliano Croce, Federico Gavazzoni, and Robert C. Ready**, "Currency risk factors in a recursive multi-country economy," *Journal of Finance*, 2017.

Cole, Harold L and Maurice Obstfeld, "Commodity trade and international risk sharing: how much do financial markets matter?," *Journal of Monetary Economics*, 1991, *28* (1), 3–24.

Collin-Dufresne, Pierre, Benjamin Junge, and Anders B Trolle, "Market Structure and Transaction Costs of Index CDSs," 2016.

Corsetti, Giancarlo, Luca Dedola, and Sylvain Leduc, "International Risk Sharing And The Transmission Of Productivity Shocks," *The Review of Economic Studies*, 2008, *75*, 443–473.

Despres, Emile, Charles P. Kindleberger, and Walter S. Salant, "The dollar and world liquidity: a minority view.," *The Economist*, February 1966, *5*, 526–529.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan, "Deviations from Covered Interest Rate Parity," Working Paper, MIT 2016.

Duffie, Darrell and Arvind Krishnamurthy, "Passthrough efficiency in the Fed's new monetary policy setting," in "Designing Resilient Monetary Policy Frameworks for the Future. Federal Reserve Bank of Kansas City, Jackson Hole Symposium" 2016.

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

___ **,** ___ **, and** ___ , "Valuation in over-the-counter markets," *Review of financial Studies*, 2007, *20* (5), 1865–1900.

Dumas, Bernard, "Dynamic Equilibrium And The Real Exchange Rate In A Spatially Separated World," *Review of financial Studies*, 1992, *5*, 153–180.

Evans, Martin D. and Richard K Lyons, "Order Flow And Exchange Rate Dynamics," *Journal of political economy*, 2002, *110* (1), 170–180.

Farhi, Emmanuel and Xavier Gabaix, "Rare Disasters And Exchange Rates," *The Quarterly Journal of Economics*, 2016, *131* (1), 1–52.

___ **, Samuel Fraiberger, Xavier Gabaix, Romain Ranciere, and Adrien Verdelhan**, "Crash Risk in Currency Markets," Working Paper 2015.

Ferrari, Massimo, Jonathan Kearns, and Andreas Schrimpf, "Monetary Policy's Rising FX Impact in the Era of Ultra-Low Rates?," Working Paper, BIS 2017.

Gabaix, Xavier and Matteo Maggiori, "International liquidity and exchange rate dynamics," *The Quarterly Journal of Economics*, 2015, *130* (3), 1369–1420.

Geanakoplos, John, "The Leverage Cycle," Technical Report, Cowles Foundation 2010.

Gertler, Mark and Nobuhiro Kiyotaki, "Financial intermediation and credit policy in business cycle analysis," *Handbook of monetary economics*, 2010, *3*, 547–599, chapter 11.

Gourinchas, Pierre-Olivier, Nicolas Govillot, and Hélène Rey, "Exorbitant Privilege And Exorbitant Duty," 2010.

Hassan, Tarek A, "Country Size, Currency Unions, And International Asset Returns," *The Journal of Finance*, 2013, *68*, 2269–2308.

Hassan, Tarek A., Thomas M. Mertens, and Tony Zhang, "Currency Manipulation," Working Paper, University of Chicago 2016.

Hau, Harald and Helene Rey, "Exchange Rates, Equity Prices, And Capital Flows," *Review Of Financial Studies*, 2006, *19*, 273–317.

___ **, Peter Hoffmann, Sam Langfield, and Yannick Timmer**, "Discriminatory Pricing of Over-The-Counter FX Derivatives," 2017. Working paper.

He, Zhiguo and Arvind Krishnamurthy, "A model of capital and crises," *The Review of Economic Studies*, 2011, *79* (2), 735–777.

___ **and** ___ , "Intermediary asset pricing," *The American Economic Review*, 2013, *103* (2), 732–770.

___ **and** ___ , "A macroeconomic framework for quantifying systemic risk. Working paper," 2014.

___ **, Bryan Kelly, and Asaf Manela**, "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 2016.

Hebert, Benjamin, "Externalities as Arbitrage," Working Paper, Stanford GSB 2017.

Holmstrom, Bengt and Jean Tirole, "Financial intermediation, loanable funds, and the real sector," *the Quarterly Journal of economics*, 1997, *112* (3), 663–691.

Itskhoki, Oleg and Dmitry Mukhin, "Exchange Rate Disconnect in General Equilibrium," Working Paper 2017.

Ivashina, Victoria, David S Scharfstein, and Jeremy C Stein, "Dollar Funding And The Lending Behavior Of Global Banks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1241–1281.

Jeanne, Olivier and Andrew K Rose, "Noise Trading And Exchange Rate Regimes," *The Quarterly Journal of Economics*, 2002, *117*, 537–569.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno N. Lustig, "Foreign Safe Asset Demand for U.S. Treasurys and the Dollar," Technical Report, Stanford GSB 2018.

Jurek, Jakub. W. and Zhikai Xu, "Option-Implied Currency Risk Premia," 2014. Working paper.

Kim, Moshe, Doron Kliger, and Bent Vale, "Estimating switching costs: the case of banking," *Journal of Financial Intermediation*, 2003, *12* (1), 25–56.

Kindleberger, Charles P., *Balance-of-payments deficits and the international market for liquidity*, Princeton, NJ: Princeton University Press., 1965.

Korinek, Anton and Alp Simsek, "Liquidity trap and excessive leverage," *The American Economic Review*, 2016, *106* (3), 699–738.

Lagos, Ricardo and Guillaume Rocheteau, "Liquidity in asset markets with search frictions," *Econometrica*, 2009, *77*, 403–426.

Law, Tzuo-Hann, Dongho Song, and Amir Yaron, "Fearing the Fed: How Wall Street Reads Main Street," Working Paper, Wharton 2017.

Liao, Gordon Y., "Credit migration and covered interest rate parity," Working Paper, Harward University 2016.

Lucas, Robert E, "Interest Rates And Currency Prices In A Two-Country World," *Journal of Monetary Economics*, 1982, *10*, 335–359.

Lustig, Hanno, Nikolai Roussanov, and Adrien Verdelhan, "Common Risk Factors In Currency Markets," *Review of Financial Studies*, 2011, *24*, 3731–3777.

Maggiori, Matteo, "Financial Intermediation, International Risk Sharing, and Reserve Currencies," *fothcoming in the American Economic Review*, 2013.

Malamud, Semyon and Andreas Schrimpf, "Intermediation Markups and Monetary Policy Passthrough," Working Paper, BIS 2016.

Malherbe, Frederic and Michael McMahon, "Financial sector origins of economic growth delusion," Working Paper, London Business School 2017.

Martin, Ian, "The forward premium puzzle in a two-country world," 2012.

Mendoza, Enrique G, Vincenzo Quadrini, and Jose-Victor Rios-Rull, "Financial Integration, Financial Development, And Global Imbalances," *Journal of political Economy*, 2009, *117*, 371–419.

Moore, Michael, Vladyslav Sushko, and Andreas Schrimpf, "Downsizing FX markets: interpreting causes and implications based on the 2016 triennial survey," *BIS Quarterly Review, December 2016*, 2016.

Mueller, Philippe, Alireza Tahbaz-Salehi, and Andrea Vedolin, "Exchange Rates and Monetary Policy Uncertainty," *Journal of Finance*, 2017, *72* (3), 1213–1252.

Obstfeld, Maurice and Kenneth Rogoff, "Exchange Rate Dynamics Redux," *Journal of political economy*, 1995, *103*, 624–660.

Pavlova, Anna and Roberto Rigobon, "Asset Prices And Exchange Rates," *Review of Financial Studies*, 2007, *20*, 1139–1180.

___ **and** ___ , "The Role Of Portfolio Constraints In The International Propagation Of Shocks," *The Review of Economic Studies*, 2008, *75*, 1215–1256.

___ **and** ___ , "An Asset-Pricing View Of External Adjustment," *Journal of International Economics*, 2010, *80*, 144–156.

___ **and** ___ , "Equilibrium Portfolios And External Adjustment Under Incomplete Markets," 2012.

Petersen, Mitchell A. and Raghuram G. Rajan, "The effect of credit market competition on lending relationships," *The Quarterly Journal of Economics*, 1995, *110* (4), 407–443.

Piazzesi, Monika and Martin Schneider, "Payments, credit and asset prices. Working paper," 2016.

Powers, Thomas Y., "The commodity currency puzzle," Technical Report, Harward University 2015.

Rampini, Adriano A and S Viswanathan, "Financial intermediary capital," 2015.

Ready, Rob, Nick Roussanov, and Colin Ward, "Commodity trade and the carry trade: a tale of two countries," *Journal of Finance*, 2017.

Richmond, Robert J., "Trade network centrality and the currency carry trade," Technical Report, UCLA 2015.

Rime, Dagfinn, Andreas Schrimpf, and Olav Syrstad, "Segmented Money Markets and Covered Interest Parity Arbitrage," Working Paper, BIS 2017.

Sharpe, Steven A., "The effect of consumer switching costs on prices: A theory and its application to the bank deposit market," *Review of Industrial Organization*, 1997, *12* (1), 79–94.

Stockman, Alan C. and Linda L. Tesar, "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements," *American Economic Review*, 1995, *85* (1), 168– 185.

Verdelhan, Adrien, "A Habit-Based Explanation Of The Exchange Rate Risk Premium," *The Journal of Finance*, 2010, *65*, 123–146.

Wu, Jing Cynthia and Ji Zhang, "Global Effective Lower Bound and Unconventional Monetary Policy," Technical Report, Chicago Booth 2018.

Previous volumes in this series

All volumes are available on our website www.bis.org.