A Quantitative Model of Limited Arbitrage in Currency Markets: Theory and Estimation

Ran Shi *

November, 2021

Link to latest version

Abstract

I develop and estimate a limits-to-arbitrage model to quantify the effects of financial constraints, arbitrage capital, and hedging demands on asset prices and their deviations from frictionless benchmarks. Using foreign exchange derivatives market data, I find that varying financial constraints and hedging demands contribute to 46 and 35 percent variation in the deviations from covered interest parity of one-year maturities. While arbitrage capital fluctuation explains the remaining 19 percent of variation on average, it periodically stabilizes prices when the other two forces exert disproportionately large impacts. The model features a general form of financial constraints and produces a nonparametric arbitrage profit function. I unveil the shapes and dynamics of financial constraints from estimates of this function.

*r.shi1@lse.ac.uk, Department of Finance, London School of Economics. I thank Vicente Cuñat, Dirk Jenter, Christian Julliard, Péter Kondor, Dong Lou, Igor Makarov, Martin Oehmke, Christopher Polk, and Walker Ray for their comments. I am greatly indebted to Ian Martin, Dimitri Vayanos, and Kathy Yuan for their advice and support.

Modern finance theory and practice build heavily on the assumption of no arbitrage. One of the textbook no-arbitrage conditions is covered interest rate parity (CIP): risk-free rates are the same for all countries after exchange rate risk is fully hedged. Before the 2008 global financial crisis, this condition broadly holds in the data.[1] After the crisis, significant and persistent CIP violations have emerged for all major currency pairs.[2] A failure of the no-arbitrage assumption has become a new normal in one of the largest financial markets in the world.

Existing limits-to-arbitrage theory provides guidance to understanding this phenomenon. First, hedging demand imbalances in the foreign exchange (FX) forwards and swaps markets can cause "price pressures", misaligning forward premiums or currency swap rates.[3] Second, arbitrageurs such as trading desks of global FX dealer banks are subject to constraints due to agency concerns or regulatory requirements.[4] They do not have the insatiable appetites to "arbitrage away" the deviations. Third, the limited amount of arbitrage capital may also forbid arbitrage positions that are large enough for eliminating price misalignments.[5]

With all the valuable theoretical perspectives, we still do not understand how quantitatively important each economic force is for CIP deviations. More broadly speaking, existing limits-to-arbitrage models generally fall short in their potential to be directly mapped to data and offer quantitative answers.[6]

This paper aims to bridge the gap. I develop a quantitative model that encapsulates the three economic forces driving CIP deviations in a parsimonious manner. Equilibrium outcomes of the model

[1] Frenkel and Levich (1975, 1977) attribute CIP deviations to transactions costs. Taylor (1987) confirms the CIP condition using high-frequency data within time windows of approximately one minute. Working with tick-by-tick data, Akram, Rime, and Sarno (2008, 2009) find that most profitable deviations last less than five minute and the CIP condition holds on average in their sample period from February 13 to September 30, 2004.

[2] Baba and Packer (2009) analyze large CIP violations during the global financial crisis. Ivashina, Scharfstein, and Stein (2015) study short-maturity CIP deviations during the Eurozone sovereign crisis, emphasizing their role as a barometer of wholesale dollar funding conditions. Du, Tepper, and Verdelhan (2018) establish the new post-crisis benchmark of CIP deviations around 25 basis points and investigate the causes.

[3] In these markets, examples for the causes of demand imbalances include i. hedging demands of currency carry traders (Du, Tepper, and Verdelhan (2018) offer suggestive evidence based on the association between cross-sectional variation in CIP deviations and average interest rate differentials); ii. financial institutions' FX risk management practices (Puriya and Bräuning (2021) identify this driving force for short-maturity FX forward contracts). Price impacts of demand shocks have also been investigated in markets of commodity futures by Acharya, Lochstoer, and Ramadorai (2013), options by Gârleanu, Pedersen, and Poteshman (2008), long-term interest rate swaps by Klingler and Sundaresan (2019), government bonds by Greenwood and Vayanos (2010), public equities by Coval and Stafford (2007) and Lou (2012).

[4] The literature on financial constraints is enormous. In the field of international finance, see Gabaix and Maggiori (2015) for their impacts on spot exchange rates. In financial economics, Gârleanu and Pedersen (2011) and Gromb and Vayanos (2018) are examples of theories examining violations of the law of one price in light of margin or collateral constraints. Andersen, Duffie, and Song (2019) explain CIP deviations in light of debt-overhang costs to equity holders of derivatives dealers.

[5] In models such as Kyle and Xiong (2001) and Kondor and Vayanos (2019), aggregate arbitrage capital endogenously creates risk-aversion dynamics, inducing commonality in asset prices in response to arbitrage capital fluctuations.

[6] Vayanos and Vila (2021) is an example of quantitative limits-to-arbitrage models for government bond markets, calibrated to predictive regression coefficients. Jermann (2020) presents and calibrates a model featuring holding costs for long-term bond to explain negative swap spreads after the financial crisis.

provide two identifying conditions for model estimation. I propose a variance decomposition scheme using the estimated model by computing counterfactual CIP deviations after holding different model ingredients constant. According to my decomposition, on average, 46 percent of variation in one-year CIP deviations of G6 currencies against the US dollar is due to changing financial constraints.[7] Demands for dollars in forward markets due to FX risk management practices of exporters and global bond investors (hedging demands) explain another 35 percent. Fluctuations in arbitrageurs' capital account for the remaining 19 percent.

Four sets of empirical findings emerge from analyzing the estimated model. First, the importance of financial constraints and hedging demands in explaining CIP deviations varies across currency pairs. Hedging demands account for 56 percent variation in the Canadian dollar CIP deviations, but less than 30 percent in the context of yen and euro. Varying financial constraints fill the vacancy left by hedging demands for these two currencies, explaining almost 60 percent variation in their basis against the dollar.

Second, arbitrageurs' capital plays a unique role in influencing the deviations. In 2009-2019, it contributes to a limited fraction of variation in CIP deviations (19 percent on average, as pointed out earlier). However, it can stabilize the basis during periods of significant variation in financial constraints or hedging demands. According to the variance decomposition results, shutting down hedging demand or financial friction variation always reduces variation in CIP deviations. However, holding arbitrageurs' capital constant can increase fluctuations in the currency basis for certain periods. During these periods, arbitrage capital counterbalances the other two forces and dampens variation in CIP deviations. For example, in 2013-2014, the one-year Canadian dollar basis is overwhelmingly driven by hedging demands. If arbitrage capital remains constant, (counterfactual) variation in one-year Canadian dollar CIP deviations would double.

Third, *shapes* of financial constraints change dramatically before and after 2014. These shapes can help understand arbitrageurs' internal capital allocation decisions in response to regulatory reforms. In 2009-2013, arbitrageurs can build CIP arbitrage positions that are at least four times larger than their equity capital without significantly downsizing other investment positions. However, in 2014-2019, building the same size of arbitrage positions will force the arbitrageurs to decrease standard investment positions by 40 percent. More importantly, a hard leverage cap of around seven (times the equity capital) emerges for the same period. This finding appears to be consistent with the fact that the supplementary leverage ratio (SLR) requirement was finalized in the third quarter of 2014. Overall, the shape of constraints after 2014 can be interpreted as a risk-weighted capital requirement plus a hard leverage ratio requirement.

Finally, bilateral net exports and net foreign security purchases are dominant forces explaining

[7] Throughout the paper, the term "G6 currencies" refers to the euro (EUR), yen (JPY), pound (GBP), Canadian dollar (CAD), Australian dollar (AUD), and Swiss Franc (CHF).

currency hedging demands.[8] For security purchases, all impacts on hedging demands come from net bond purchases (as oppose to equity). The estimated demand functions suggest that *forward* dollar demands increase in the US net exports and net bond purchases. In other words, exporters and investors holding foreign bonds hedge more when they need to repatriate more future incomes denominated in foreign currencies.

I now describe the model to provide intuitions on its mechanism and estimation. In the model, competitive arbitrageurs trade with hedgers; trading determines equilibrium arbitrage yields such as CIP deviations. Hedgers exchange specific currencies for dollars forward. I build an optimizing foundation for this forward dollar demand in a two-country setting. Hedgers from each country are subject to endowment shocks denominated in foreign currencies. They manage their exchange rate risk using forward contracts. Differences in their foreign endowments create hedging demand imbalances described by the demand functions specified in the model.

Competitive arbitrageurs maximize (additive) discounted log utilities over their lifetime consumption stream.[9] They profit from multiple arbitrage opportunities by absorbing demand imbalances in FX derivatives markets for different currency pairs. In the meantime, they have access to standard investment opportunities: one risk-free and one risky asset. Arbitrage limits arise from financial constraints on both their arbitrage positions and investment positions. The constraints induce a trade-off between deploying capital to their standard investment business and diverting the resource to arbitrage activities. Sizable arbitrage positions come at the cost of cutting back routine investment positions.

The model features an agnostic view regarding the specific form of financial constraints.[10] This setup encompasses standard specifications of frictions such as the margin requirements, leverage ratio requirements, Value-at-Risk rules, transaction costs, and credit/debt/funding valuation adjustments. This generality in modeling choice allows me to reach robust theoretical conclusions and perform nonparametric estimation of the constraints. This approach is particularly helpful given the sophisticated nature of post-crisis financial regulations. In this new era, numerous regulatory constraints exist and many of them can be binding at the same time.

Even without the dedication to a certain form of financial constraints, the model still yields strong predictions. Most importantly, the model argues that the absolute values of present arbitrage yields

[8]Net foreign security purchases are the difference between domestic (US) residents' purchases of foreign securities and foreign residents' purchases of domestic securities. As a clarification, this measure does not necessarily represent portfolio flows as "round-trip" trades can occur between international investors from different countries.

[9]Simplified version of the model lasts two periods; the full model comes with infinite time horizon. I also extend the theory for general constant relative risk-aversion (CRRA) utilities.

[10]In the model, arbitrageurs solve dynamic consumption and portfolio choice problems with one risk-free asset, one risky asset, and multiple riskless arbitrage opportunities, under general position constraints: a bounded, closed, and convex set. The specification of constraints is the same as Cvitanić and Karatzas (1992), who solve the Merton model of one risk-free asset and multiple risky assets under general position constraints. One theoretical contribution is that I characterize arbitrageurs' optimality conditions using a more accessible primal-dual approach.

(e.g., CIP deviations) predict future returns on arbitrageurs' capital,[11] and this predictive relationship is convex. This increasing and convex function (of capital returns in response to arbitrage yields) reveals the form of financial constraints. For example, under margin requirements, the function equals zero when deviations are small and increases linearly after a threshold. However, with Value-at-Risk rules, arbitrageurs' capital returns smoothly respond to all levels of deviations regardless how small they are. The function is defined in the same way as the profit function in standard production theories, thus named the *arbitrage profit function*. A more convex arbitrage profit function implies a higher hurdle for arbitrageurs to covert CIP deviations into sizable arbitrage profits.

Empirical evidence supports the prediction. Average CIP deviations among G6 currencies predict monthly and quarterly returns on arbitrageurs' capital, after controlling for common time-series return predictors. One basis point increase in the average deviations forecasts at least two percentage point increase in the (annualized) returns. Two statistical tests, one parametric and another semi-parametric confirm that the predictive relationship is convex.

Model estimation takes two steps, both relying on equilibrium outcomes of the model. The first step is to estimate the arbitrage profit function, which characterizes the increasing and convex response of arbitrageurs' capital returns to CIP deviations. The model allows this function to change across time, reflecting varying stringency of financial constraints. I develop a new statistical procedure to estimate both the functional form and time-series variation of arbitrage profit functions in the model. The functional form reveals how the (binding) financial constraints look like collectively; the time-series variation describes dynamics of the constraints. I compute equilibrium arbitrage positions using the these estimates, which can be inferred from market prices once the arbitrage profit function is known.

The second step is estimating parameters in hedging demand functions. In equilibrium, CIP deviations in the model are such that arbitrage positions equal hedging demands. Plugging-in the inferred arbitrage positions enable demand estimation without using position data in FX derivatives markets. To resolve the endogeneity concerns about CIP deviations and latent demands (unobservable drivers of hedging demands), I construct an instrumental variable estimator for demand elasticities. The instruments for the deviations of one currency are observable hedging demand drivers of *other* currencies. The identification strategy is motivated by the fact that FX arbitrageurs such as global dealer banks can profit from multiple currency basis simultaneously. For a specific currency, hedging demand drivers of other currencies shift arbitrage profits. Arbitrage positions exploiting the CIP deviations of this particular currency change accordingly. The instruments effectively become "supply shifters" (if we interpret arbitrageurs as suppliers of "arbitrage services") that are uncorrelated with latent demands.

[11] Though CIP basis can be positive (e.g., Australian dollars against US dollars) or negative (e.g., euro or yen against US dollars), arbitrageurs can always profit from it by properly switching legs of their positions. Thus, absolute values of the deviations contribute to arbitrage profits. I will stop mentioning "the absolute values" later on in the introduction for the ease of exposition.

Broader contribution of this paper is twofold. Typically, there is a separation of theory and empirics in the limits-to-arbitrage literature. I aim to partially bridge the gap by building a model that synthesizes necessary ingredients in the existing theory and maps directly to the data to quantify the economic forces at work. The backbone of my model is close to Gabaix and Maggiori (2015), who study real imbalances in the currency spot markets absorbed by "financiers" facing commitment problems (which translate into quadratic position limits). My model focuses on hedging demand imbalances in FX forwards and swaps markets. It specifies financial frictions in a general format. I further quantify demand imbalances and financial frictions to explain empirical patterns and facilitate counterfactual exercises.

The second contribution is methodological: the estimation framework can be applied to other violations of the no-arbitrage condition in today's financial markets. I demonstrate how to back out the financial constraints and arbitrage positions from arbitrageurs' capital returns and arbitrage yields. The flexible nonparametric arbitrage profit function approach is particularly useful in light of the numerous regulatory reforms after 2008. Unlike the demand system approach to asset pricing (Koijen and Yogo, 2019), my methodology for estimating demand function parameters does not rely on position data (though high-quality position data can help discipline my estimation), but instead draws inferences using price data based on arbitrageurs' optimality conditions.

The paper is structured as follows. Section 1 briefly reviews the definition and measurements of CIP deviations to provide additional backgrounds. Section 2 presents a simplified version of the model to illustrate key intuitions. Section 3 introduces the full model and characterizes its equilibrium outcomes. Section 4 describes additional data and measurements, tests the model's main prediction, enriches the model to map it to data, and describes estimation methodologies. Section 5 performs quantitative exercises using the estimated model. Section 6 concludes. All proofs are in the Appendix.

1 CIP deviations and their measures

At time t, the CIP deviation for currency i of maturity τ is b such that the following equation holds:

$$\exp\left(r^{\$}_{t\to(t+\tau)}\tau\right) = \exp\left(r^{i}_{t\to(t+\tau)}\tau + b\tau\right)\frac{F_{t\to(t+\tau)}}{E_t}, \tag{1}$$

where $r^{\$}_{t\to(t+\tau)}$ and $r^{i}_{t\to(t+\tau)}$ represent risk-free rates of the US dollar and currency i from time t to $(t+\tau)$; $F_{t\to(t+\tau)}$ is the forward price of currency i in dollars maturing at time $(t+\tau)$; E_t is the spot price.

Following Du, Tepper, and Verdelhan (2018), I focus on two measures of CIP deviations using different derivative contracts: FX forwards and cross-currency basis swaps (currency or basis swaps in short). From FX forward contracts, observable currency forward prices $F_{t\to(t+\tau)}$ (thus observable forward premiums $F_{t\to(t+\tau)}/E_t$) can be plugged in to the equation above. The two risk-free rates $r_t^{\$}$ and r_t^i are commonly measured by over-night index swap (OIS) rates for different countries. I call CIP deviations calculated from equation (1) using these variables the forward-OIS bases.

A more direct measure of CIP deviations comes from the currency swap contracts. In a currency swap contract, two parties (namely Alice and Bob) exchange currencies at spot rates upfront and pay each other back effectively with floating rate bonds. Specifically, Alice, receiving £100 from Bob initially, will pay Bob back with (cashflows of) a pound floating rate bond (face value = £100); Bob, receiving \$135 (let $E_t = 1.35$ be the GBP/USD spot rate) from Alice at beginning of the contract, will pay Alice back with a dollar floating rate bond (face value = \$135). Currency swap contracts quote b such that Bob pays the the dollar floating rates $\{r^{\$}_{t\to(t+\Delta t)}, r^{\$}_{(t+\Delta t)\to(t+2\Delta t)}, \dots\}$, and Alice pays *adjusted* pound floating rates $\{r^{£}_{t\to(t+\Delta t)}, r^{£}_{(t+\Delta t)\to(t+2\Delta t)}, \dots\} + b$. The payments are usually made on a quarterly basis (i.e., $\Delta t = 0.25$). Back to equation (1), we can interpret this quoted currency swap rate as CIP deviations defined through treating $r^{\$}_{t\to(t+\tau)}$ and $r^{i}_{t\to(t+\tau)}$ as interest rate swap rates (swapping the two floating rates). Du, Tepper, and Verdelhan (2018) and Augustin, Chernov, Schmid, and Song (2020) describe detailed trading arrangements justifying this conclusion.

Throughout this paper, I focus on one-year currency swap rates and use forward-OIS implied one-year CIP deviations for validation. At this maturity, both the FX forwards and currency swaps have high trading volumes and low bid-ask spreads. I collect the FX forward/spot prices, OIS rates, and currency swap rates from Bloomberg. Table 1 reports summary statistics of the two deviation measures; Figure A1 in the Appendix plots these two measures for G6 currencies. Overall, currency swap rates offer more conservative and less volatile measures of CIP deviations compared with the forward-OIS bases at the one-year maturity.

2 A simple model in a two-period deterministic economy

To begin with, I present a simple model with no uncertainty to illustrate key insights of the model. As a preview, the model features an agnostic view about the specific constraints arbitrageurs face, predicts a convex relationship between arbitrageurs' investment return and their arbitrage profit (from CIP deviations), and determines CIP deviations as tractable equilibrium outcomes.

The economy lasts for two periods: today and tomorrow.[12] There are two types of agents: arbitrageurs and hedgers. Each type composes an identical continuum of measure one.

Arbitrageurs. Each arbitrageur is endowed with initial capital k today. They choose their consumptions and derive utilities from them as follows:

$$\log(y) + \frac{1}{1+\rho}\log(y'). \tag{2}$$

[12] Notation-wise, variables tomorrow come with prime superscripts.

The subjective time discount rate $1/(1+\rho)$ belongs to the interval $(0, 1)$, i.e., $\rho > 0$; consumptions are y today and y' tomorrow.

The arbitrageurs can invest in a risk-free asset earning a net return r ($r > 0$), or simply store their capital safely with zero net return. I assume that the amount of capital stored cannot be negative.[13] Arbitrageurs can also profit from a riskless arbitrage opportunity, yielding b per unit of position they enter. For currency markets, we can interpret b as CIP deviations, which is either positive (e.g., Australian dollars) or negative (e.g., yen).

By consuming y today, arbitrageurs are saving (or equivalently, investing) $s = (k - y)$ amount of capital to fund their future consumption. Denote by π_0 and π the "weights" of their investment positions in the risk-free asset and the arbitrage opportunity, their absolute positions are $\pi_0 s$ and πs accordingly. To earn the risk-free rate of return, capital is needed: $(1+r)\pi_0 s$ units of capital next period come at a cost of $\pi_0 s$ today. In comparison, harvesting the arbitrage profits $\pi s b$ next period requires no capital today. As a result, the arbitrageurs' capital next period k' is given by

$$k' = s + \pi_0 s(1+r) - \pi_0 s + \pi s b - 0 = s\left[1 + \pi_0 r + \pi b\right]. \tag{3}$$

According to equation (3), arbitrageurs store $(1-\pi_0)s$ units of capital (after investing $\pi_0 s$ in the risk-free asset). As the economy lasts for only two periods, arbitrageurs consume all their capital tomorrow, i.e., $y' = k'$.

Replacing y' in problem (2) with $s\left[1 + \pi_0 r + \pi b\right]$, we can see that arbitrageurs are solving two separate problems:

$$\underset{y,\, s=k-y}{\text{maximize}} \quad \log(y) + \frac{1}{1+\rho}\log(s) \quad \text{and} \quad \underset{\pi_0 \leq 1,\, \pi \in \mathbb{R}}{\text{maximize}} \quad \log(1 + \pi_0 r + \pi b). \tag{4}$$

The restriction $\pi_0 \leq 1$ appearing in the second problem is due to the assumption that capital stored is nonnegative, that is, $(1-\pi_0)s \geq 0$.

Hedgers. Hedgers use currency forwards or swaps to manage their foreign exchange exposures. Under the context of CIP deviations, I assume that their (aggregate) demand for selling foreign currencies (say, pounds) in exchange for dollars *in forward markets* is

$$q(b) = \gamma_0 - \gamma b, \quad \gamma > 0. \tag{5}$$

These *forward dollar demands* are downward sloping with regard to the CIP deviation b. This is

[13] This claim rules the possibility that arbitrageurs can raise fund by paying a gross interest rate of one (the storage yield), for this itself leads to another riskless arbitrage within the model: borrowing at cost one, investing in the risk-free asset yielding $(1+r)$. The storage technology is needed in the model because, after introducing arbitrage limits later in the paper, arbitrageurs need to devote capital to arbitrage positions. The capital buttressing their arbitrage activities is "stored" in the sense that they cannot generate a return as high as r. We can treat the zero storage yield here as a normalization.

because, according to equation (1), a smaller b for the pound is equivalent to higher forward price of pounds against dollars. It propels hedgers' willingness to sell pounds for dollars forward, creating higher forward dollar demands.[14]

The equilibrium arbitrage yield. Arbitrageurs "take the opposite side" against hedgers' demands: their arbitrage positions are effectively *forward dollar supplies*. When hedgers sell pounds for dollars forward (positive forward dollar demand, $q > 0$), pressing GBP/USD forward price to drop below the no-arbitrage benchmark, a positive CIP deviation emerges.[15] In response, arbitrageurs can take advantage of this opportunity by borrowing pounds (yielding $-r^{£}$), swapping pounds to dollars (yielding $r^{£} + b - r^{\$}$), and lending dollars (yielding $+r^{\$}$). They supply dollars in the currency forward markets because of the need to payback the dollars they received at the beginning of the swap contract. A more simplistic view is that with $b > 0$, the forward price of GBP/USD is relatively low, arbitrageurs tend to offer (supply) dollars to buy pounds forward. Their total supply of dollars $\pi(b)s$ is positive,[16] in which $\pi(b)$ solves (4) for a given b. The equilibrium deviations solve the following equation:

$$\pi(b)s = q(b). \tag{6}$$

Of note, all the analysis goes through in the same way when there is a negative forward dollar demand, i.e., $q < 0$.[17]

The equilibrium b that solves (6) is such that

$$b = \frac{\gamma_0}{\gamma + \pi(b)/b \times s}. \tag{7}$$

[14]Appendix C provides an optimization foundation to the reduced form specification (5). I build a two-country currency-risk hedging model, in which US hedgers manage their currency exposures through selling pounds and UK hedgers conduct the opposite trade in pound-dollar forward market. Their hedging needs do not necessarily cancel out, which give rise to the (net) demands specified in (5), representing hedging demand *imbalances* in currency markets.

The hedging demand $q(b)$ can take either positive or negative signs. According to the micro-foundation in Appendix C, US hedgers offer forward pounds for dollars while UK hedgers seek opposite trades. When the US hedgers' demand exceeds its UK counterpart, $q(b)$ is positive. The net effect is a positive demand for forward dollars. This demand becomes negative when the UK hedgers hedge more. In other words, a negative $q(b)$ can be interpreted as the net demand for foreign currencies (by selling dollars forward).

When the demand q is negative, hedgers are selling forward dollars in exchange for pounds, causing negative forward dollar demands. As b becomes smaller (thus a higher forward GBP/USD price F or, to put it differently, a lower dollar forward price), hedgers tend to sell less dollar: q still increases as b decreases.

[15]Without frictions, arbitrageurs absorb hedging demands imbalances "with ease" and equilibrium deviations always equal zero. This ideal outcome As we will show later in Proposition 2, when there are arbitrage limits, a positive forward dollar demand, $q > 0$, is equivalent to both $\gamma_0 > 0$ and the equilibrium CIP deviations b^* to be $0 < b^* \leq \gamma_0/\gamma$.

[16]I provide rigorous theoretical arguments for this through Lemma A.1 in the Appendix.

[17]With negative forward dollar demand, i.e., $q < 0$, Proposition 2 presented later commands a negative b. Arbitrageurs will borrow dollars (yielding $-r^{\$}$), swap dollars for pounds (yielding $r^{\$} - r^{£} - b$), and lend pounds (yielding $r^{£}$). Their optimal arbitrage position $\pi(b)s$ is negative, which implies a negative supply of forward dollars (demanding dollars forward). This negative supply is due to the fact that arbitrageurs will receive forward dollars and return pounds at the end of their swap contracts.

The frictionless benchmark. Without friction, the second maximization problem in (4) commands $\pi_0 = 1$ and $|\pi(b)| \to \infty$ whenever $|b| > 0$. This implies that $\pi(b)/b \to \infty$ for any b around a neighborhood of zero. According to equation (7), the equilibrium deviation is zero. Absence of arbitrage in the model is a result of "hyper-elastic" arbitrage positions in response to arbitrage yields.

Limits to arbitrage. Arbitrage limits that cause CIP deviations must prevent $\pi(b)/b$ from going to infinity whenever b deviates from zero. I assume that they arise from the following position constraint on π_0 and π:

$$(\pi_0, \pi) \in \mathcal{C}, \tag{8}$$

in which $\mathcal{C}$ is a subset of $(-\infty, 1] \times \mathbb{R}$ (domains defined in the second problem of (4)), outside of which the combination of π_0 and π becomes infeasible. Under this assumption, the second problem of (4) is equivalent to maximize $(\pi_0 r + \pi b)$ subject to the condition that $(\pi_0, \pi) \in \mathcal{C}$. The outcome from solving this problem represents the optimal return on investment for the arbitrageurs, denoted by

$$S_{\mathcal{C}}(r, b) = \sup_{(\pi_0, \pi) \in \mathcal{C}} \{\pi_0 r + \pi b\}. \tag{9}$$

$S_{\mathcal{C}}$ is often named the *support function* of the set $\mathcal{C}$. It works the same way as the profit function in the standard production theory, when the set $\mathcal{C}$ is a production set (Mas-Colell, Whinston, and Green, 1995, Chapter 5.B-5.C, p. 128-143). We can call this function the *arbitrage profit function*. As I will illustrate soon, this function defines the optimal investment return per unit of capital for arbitrageurs.

The trade-off arbitrageurs face is fully characterized by the position constraint. When they extend their positions to take advantage of an arbitrage opportunity, they have to cut positions on the risk-free asset (i.e., put a fraction $(1 - \pi_0)$ of their capital inefficiently in storage). Facing this trade-off, arbitrageurs optimally allocate their capital such that they enjoy a (net) return of $S_{\mathcal{C}}(r, b)$ per unit of savings. As a result, in equilibrium, $k' = s[1 + S_{\mathcal{C}}(r, b)]$. Now I enumerate four assumptions about the constraint $\mathcal{C}$ and one assumption about the arbitrageurs' positions.

Assumption 1. *$\mathcal{C}$ is a subset of* $[0, 1] \times \mathbb{R}$.

The assumption that $\pi_0 \leq 1$ reiterates $(1 - \pi_0)s \geq 0$, that is, "negative storage" is not allowed – arbitrageurs cannot borrow money at a zero net interest rate. Assuming $\pi_0 \geq 0$ forbids arbitrageurs from borrowing at the rate r and then storing the proceeds (to further enlarge their arbitrage positions after exhausting all their initial capital k).

Assumption 2. *$\mathcal{C}$ is bounded, closed, and convex.*

The boundedness assumption is straightforward, under which $S_{\mathcal{C}}(r, b) < \infty$. Closeness of the set $\mathcal{C}$ means that for any unattainable combination (π_0, π) (falling in the complement of $\mathcal{C}$, an open set), a small neighbor of this combined position is still infeasible for the arbitrageurs to take on: unachievable positions do not suddenly become feasible. Convexity of $\mathcal{C}$ means that convex combinations of feasible position pairs are still available to the arbitrageurs.

Assumption 3. *"Going all in" on the risk-free asset is allowed for the arbitrageurs, that is,* $(1, 0) \in \mathcal{C}$.

From this assumption, we have $S_{\mathcal{C}}(r, b) \geq r$, the optimal return per unit of savings invested is at least r. Thus, taking advantage of arbitrage opportunities benefits the arbitrageurs, although this activity may require inefficient storage of arbitrage capital.

Assumption 4. *When the arbitrage yield b equals zero, the arbitrage position is zero, that is,* $\pi(0) = 0$.

This is a behavioral assumption about the arbitrageurs. When $b = 0$, the total arbitrage profit is always zero and arbitrage positions π can take any value. Assumption 4 restricts the positions to zero. We can interpret this assumption as arbitrageurs regard the riskless arbitrage opportunity as simply nonexistent whenever its yield equals zero.

Three examples illustrate the set $\mathcal{C}$ under these assumptions and characterize arbitrageurs' optimal choices.

Example 1 (Margin requirements). Margin requirements as highlighted in Gârleanu and Pedersen (2011) can prevent arbitrageurs from building up a large derivative position to "arbitrage away" the opportunities such as CIP deviations. Following their convention (of symmetric margins[18]), I let $\mathcal{C}$ be $\{0 \leq \pi_0 \leq 1, \pi \in \mathbb{R} : \pi_0 + m|\pi| \leq 1\}$. Under this specification, arbitrageurs need to post collaterals into margin accounts for their derivative positions: for one unit increase in the notional value, m units of capital are occupied, thus not available for investing in the risk-free asset.[19] With margin requirements, the arbitrage position is

$$\pi(b) = \frac{\text{sgn}(b)}{m} I_{\{|b| \geq mr\}}.$$[20]

Arbitrageurs behave in an "all-or-nothing" manner: they remain inactive when the arbitrage yield is small; otherwise, they build arbitrage positions to the fullest capacity. Panel (A) of Figure 1 shows the shape of $\mathcal{C}$ and plots $\pi(b)$.

Example 2 (Costs and adjustments). Now consider the case that a total arbitrage position of value πs will incur a cost or adjustment of $C(\pi s, s)$.[21] Adopting a standard specification for adjustment cost functions in investment theory (e.g., Hayashi (1982)), I assume $C(\pi s, s) = c(\pi)s$, that

[18]Extending the characterization to asymmetric margin requirements changes the constrain to $\pi_0 + m^+\pi^+ + m^-\pi^- \leq 1$ where m^+ and m^- apply to long (π^+) and short (π^-) legs of derivative contracts respectively.

[19]An implicit assumption here is that capital posted in the margin account are "stored" using the one-to-one storage technology. In practice, money in the margin account earns interest. Then we could interpret this implicit assumption as a normalization argument, that is, all prices r and b will be normalized by the margin account compensation rate.

[20]The signum function $\text{sgn}(b)$ equals -1 when $b < 0$, 0 when $b = 0$, and 1 when $b > 0$.

[21]In a standard (I, K) type of investment theory presented as early as by Lucas (1967), adjustment costs are relate to

is, the cost function is (positively) homogeneous of degree one. As a result, the budget constraint (3) is now $k' = s\left[1 + \pi_0 r + \pi b - c(\pi)\right]$. Define $\hat{\pi}_0 = \pi_0 - c(\pi)/r$, the optimization problem of (9) becomes maximizing $\hat{\pi}_0 r + \pi b$ subject to the condition that $\mathcal{C} = \{0 \le \hat{\pi}_0 \le 1,\, \pi \in \mathbb{R} : \hat{\pi}_0 + c(\pi)/r \le 1\}$. To make it more specific, I let the function $c(\pi)$ be quadratic with regard to $|\pi|$, that is, $c(\pi) = G|\pi| + (1/2)g\pi^2$ ($G \ge 0$, $g \ge 0$). I present $\mathcal{C}$ and the optimal arbitrage position $\pi(b)$ in Panel (B) of Figure 1 under this specification. Similar to margin requirements, there is still a region of inaction for the arbitrageurs: they do not respond when $|b| < G$. However, when arbitrage yields are moderately large, that is, when $|b|$ is greater than G but still smaller than $\sqrt{G^2 + 2gr}$, arbitragers gradually increase their positions, until exhausting all their capital. Clearly, if $g = 0$, meaning that the quadratic term $(1/2)g\pi^2$ disappears from the cost function $c(\pi)$, this example collapses to the one under margin requirements where $m = G/r$.

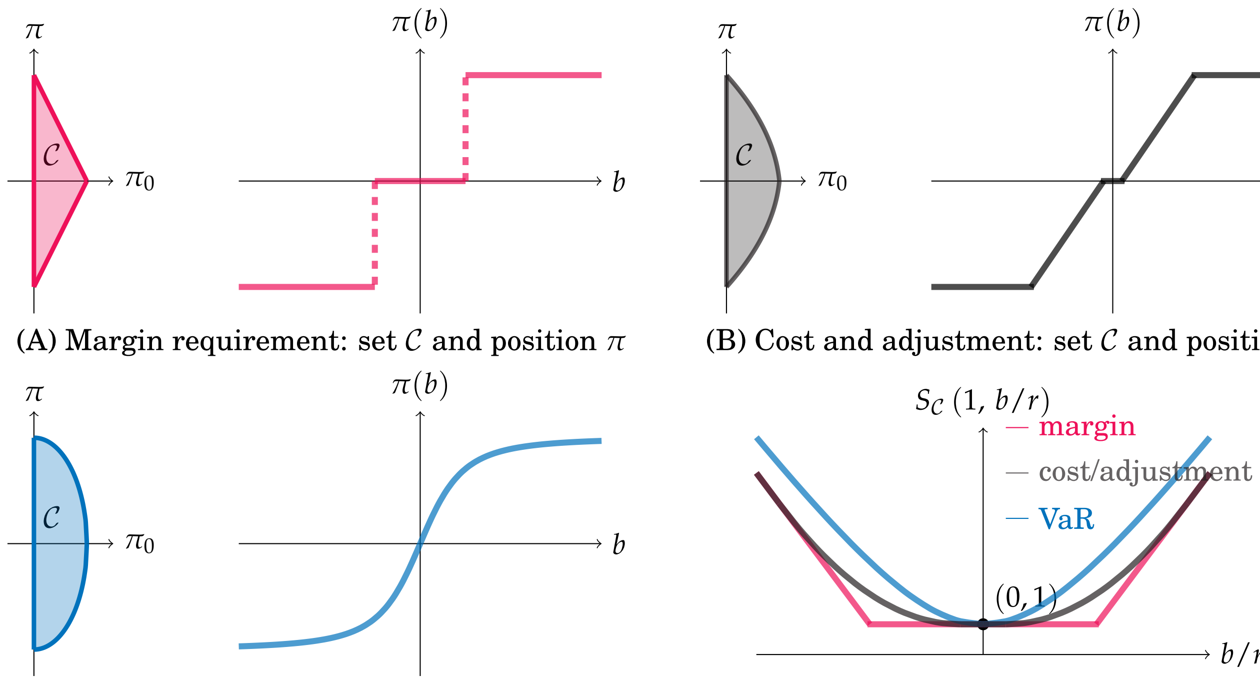
Example 3 (Value-at-Risk constraints). Another family of constraints arbitrageurs can face result from Value-at-Risk (VaR) calculations as highlighted by Adrian and Shin (2014) in the study of bank leverage. Under this rule, arbitrageurs need enough equity capital to cover their VaR_α, defined as $\inf\{V > 0 : \mathbb{P}\left[\text{change in asset value} \le -V\right] \le 1 - \alpha\}$, based on a pre-specified small threshold α. Arbitrageurs adjust their investment positions to abide by the rule. As an illustration, I consider a simple Gaussian VaR setting, under which changes in r and b are both normal; for simplicity, I further assume that these changes are independent. In summary, $\Delta r \sim \mathcal{N}(\mu_{\Delta r}, \sigma_{\Delta r})$, $\Delta b \sim \mathcal{N}(\mu_{\Delta b}, \sigma_{\Delta b})$ and $\Delta r \perp \Delta b$. Under this setting, $\mathrm{VaR}_\alpha = z_{(1-\alpha)}\sqrt{\pi_0^2 s^2 \sigma_{\Delta r}^2 + \pi^2 s^2 \sigma_{\Delta b}^2}$, where $\sigma_{\Delta r}$ and $\sigma_{\Delta b}$ can be calibrated from historical data, $z_{(1-\alpha)}$ is the $[100(1-\alpha)]$th percentile of the standard normal distribution. Thus the VaR constraint $\mathrm{VaR}_\alpha \le s$ yields $z_{(1-\alpha)}^2\left(\sigma_{\Delta r}^2 \pi_0^2 + \sigma_{\Delta b}^2 \pi^2\right) \le 1$. As a normalization, we can let $z_{(1-\alpha)}^2 \sigma_{\Delta r}^2 = 1$ and define $v = \sigma_{\Delta b}/\sigma_{\Delta r}$, then the set $\mathcal{C}$ becomes $\left\{0 \le \pi_0 \le 1,\, \pi \in \mathbb{R} : \pi_0^2 + v^2\pi^2 \le 1\right\}$. Under this VaR constraint, arbitrageurs choose their arbitrage positions

$$\pi(b) = \frac{b}{v\sqrt{r^2 v^2 + b^2}}.$$

This setting features smooth arbitrage responses to the magnitude of arbitrage yeilds, in sharp contrast to the outcomes under margin requirements. Panel (C) of Figure 1 illustrates the set $\mathcal{C}$ as well as the function $\pi(b)$.

Panel (D) of Figure 1 compares the arbitrage profit function $S_{\mathcal{C}}$ for the three examples. Of note, $S_{\mathcal{C}}$ is positively homogeneous of degree one (e.g., Molchanov and Molinari (2018, p. 75)), thus the plot shows $S_{\mathcal{C}}(1, b/r)$ as a function of b/r for cleaner demonstration. $S_{\mathcal{C}}(1, b/r)$ reaches its minimum

both the investment I (πs here) and the capital stock K (s here). This is because the relative size of I given K may help determine the cost. The costs or adjustments may be due to funding value adjustments as demonstrated in Andersen, Duffie, and Song (2019), which is an implicit debt-overhang cost to equity holders. Arbitrageurs' effective funding costs and (opportunity) costs of collaterals for different currency pairs may also render CIP arbitrage less profitable (Augustin, Chernov, Schmid, and Song, 2020). And, as many may argue, counterparty credit risk adjustments may also plague the seemingly riskless CIP arbitrage, for most currency derivatives are not centrally cleared (though unfavorable evidence provided in Du, Tepper, and Verdelhan (2018)). The cost function here can also be interpreted as (credit) risk adjustments.



(A) Margin requirement: set $\mathcal{C}$ and position π

(B) Cost and adjustment: set $\mathcal{C}$ and position π

(C) VaR constraint: set $\mathcal{C}$ and position π

(D) Arbitrage profit function

Figure 1: Examples of arbitrage constraints $\mathcal{C}$, arbitrage positions, and arbitrageurs' optimal investment returns.

value of one at $b = 0$. This is when the arbitrage opportunity does not exist, so the investment return must be r. When $|b|$ deviates from zero, $S_{\mathcal{C}}(1, b/r)$ will never decrease.

The three special cases of $\mathcal{C}$ exemplify the benefits of developing a theory *without* taking a strong stand on the form of the constraint. Different shapes of $\mathcal{C}$ lead to distinctive arbitrage responses, which translate into peculiar (in)elasticities of $\pi(b)$, the supply of forward dollars. From equation (7), equilibrium arbitrage yields thus differ. I summary theoretical results based on this agnostic view of arbitrage limits in propositions below.

Proposition 1. *The optimal behavior of arbitrageurs imposes the following equilibrium conditions:*

$$\frac{1}{1+\rho}\left(\frac{y'}{y}\right)^{-1}\left[1 + S_{\mathcal{C}}(r,\, b)\right] = 1$$

for their consumption growth (the Euler equation) and

$$\frac{1}{r}\left(\frac{k'-k}{k}\right) = \left(\frac{1}{2+\rho}\right) S_{\mathcal{C}}\left(1,\, \frac{b}{r}\right) - \left(\frac{1+\rho}{2+\rho}\right)\frac{1}{r}$$

for their capital accumulation. All else equal, the net return on arbitrageurs' capital $[(k'-k)/k]$*: i.*

increases in $|b|$;[22] *ii. is a convex function of* b.

The consumption Euler equation is standard under the log utility, in which $(1+S_{\mathcal{C}})$ acts as the return on intertemporal savings. Motivated by this equation, we can conceptualize arbitrageurs' decision problem as a two-stage one: they first optimize $\pi_0 r + \pi b$ subject to the constraint (8), which gives the optimal return $S_{\mathcal{C}}$; next, they choose their consumption plan y (and thus s, k', and y') according to the Euler equation, taking $S_{\mathcal{C}}$ as given. Arbitrage limits only prevent them from responding insatiably to arbitrage profits. Their intertemporal saving behavior remains optimal under any predetermined position constraints.

Specifications of $\mathcal{C}$ directly affects how arbitrageurs' capital accumulation responds to arbitrage yields. For example, under VaR constraints, a nonzero b lifts arbitrageurs' investment return above r, regardless how small $|b|$ is. However, with margin requirements, there exists a region around zero, within which no value of b increases the arbitrageurs' investment return.

Now we turn to optimal arbitrage positions and the equilibrium arbitrage yield. The proposition below summarizes the results.

Proposition 2. *If* $\mathcal{C}$ *is such that the support function* $S_{\mathcal{C}}$ *is differentiable, the arbitrageurs' optimal arbitrage positions are*

$$\pi(b) = \frac{\partial S_{\mathcal{C}}(r, b)}{\partial b}.$$

Furthermore, if $\mathcal{C}$ *is such that* $S_{\mathcal{C}}$ *is twice differentiable, the equilibrium deviation* b^* *that solves* $\pi(b)s = q(b)$ *uniquely exists and*

i. if $\gamma_0 \geq 0$, $0 \leq b^* \leq \gamma_0/\gamma$ *and* $q(b^*) \geq 0$*; otherwise,* $\gamma_0/\gamma \leq b^* \leq 0$ *and* $q(b^*) \leq 0$*;*

ii. $|b^*|$ *decreases as the arbitrageurs' initial capital* k *increases.*

From Proposition 2, we know that optimal arbitrage positions can be derived from the arbitrage profit function $S_{\mathcal{C}}$.[23] This function $S_{\mathcal{C}}$, on the other hand, reflects how arbitrageurs' investment return responds to arbitrage yields (Proposition 1). These observations lay foundations for identifying the forward dollar supply by the arbitrageurs. If we can measure the capital return $(k'-k)/k$, a nonparametric regression of this return on the arbitrage yield (e.g., CIP deviations) reveals the $S_{\mathcal{C}}$, which in turn gives us $\pi(b)$. We will revisit this idea later in the full quantitative model in Section 3.

The sign of equilibrium deviations is determined only by the hedgers' demand $q(b)$. Negative CIP deviations indicate that there is a net demand for foreign currencies ($q(b) < 0$) while positive

[22] Strictly speaking, all increasing or decreasing statements henceforward refer to nondecreasing or nonincreasing respectively. I avoid invoking the latter terms for conceptual simplicity, disregarding mathematical rigor.

[23] At points that the partial derivative $\partial S_{\mathcal{C}}(r, b)/\partial b$ is not well-defined, indeterminacy can arise and $\pi(b)$ falls into a closed convex set, namely the *subdifferential* of $S_{\mathcal{C}}$. See Bertsekas (2009, p. 182-186) for further expositions.

deviations imply a net demand for dollars ($q(b) > 0$), in currency forwards and swaps markets. The largest possible absolute deviation in equilibrium $|b^*|$ is always less than $|\gamma_0|/\gamma$, which is the outcome when no arbitrage force exists to absorb the hedging demand imbalances. In this equilibrium, b^* is such that $q = 0$.

The log-utility assumption brings up wealth effects, thus arbitrageurs' capital k have major impacts on their absolute arbitrage positions, which equals πs.[24] Larger capital stock increases arbitrage capacity, leading to smaller arbitrage yields in equilibrium.

3 A quantitative equilibrium model of limited arbitrage

In this section, I develop a quantitative model of limited arbitrage in currency market by enriching the simple model of Section 2. The extension comes from four dimensions: (i) a risky project is now available to the arbitrageurs; (ii) multiple (instead of only one) riskless arbitrage opportunities exist; (iii) the model is dynamic in which arbitrageurs optimize their discounted life-time utility; (iv) time-varying external hedging demand exists for each arbitrage opportunity. The risky project and random hedging demands make the model stochastic. I characterize equilibrium outcomes of the model, test their predictions, and use the equilibrium conditions to estimate the model.

3.1 Model setup

Time is continuous, going from zero to infinity. As before, there are two groups of agents: arbitrageurs and hedgers, both of a continuum of mass one.

Of note, throughout the rest of the paper, I will omit the time subscripts whenever it does not cause confusion.

Arbitrageurs. Arbitrageurs maximize a utility function

$$\mathbb{E}_t \left[\int_0^\infty e^{-\rho s} \log \left(y_{t+s} \right) \, \mathrm{d}s \right], \tag{10}$$

in which $\rho > 0$ is the instantaneous time discount rate, and y_t is their rate of consumption at date t. At date 0, they are endowed with $k_0 > 0$ amount of capital.

As before, with date-t capital k_t at hand, arbitrageurs can borrow or save at a risk-free rate r_t, or safely store their capital (with zero net return). They can also profit from multiple riskless arbitrage opportunities, yielding at rate b_{it}, $i = 1, \dots, n$ ($n \geq 1$), per unit of position they build up. In currency markets, these arbitrage yields are CIP deviations for different currencies.

Arbitrageurs now have access to a risky project, the net return of which follows a diffusion process

[24] As shown in the Appendix, in equilibrium, arbitrageurs' savings s is proportional to their initial capital endowment k, due to the log-utility assumption.

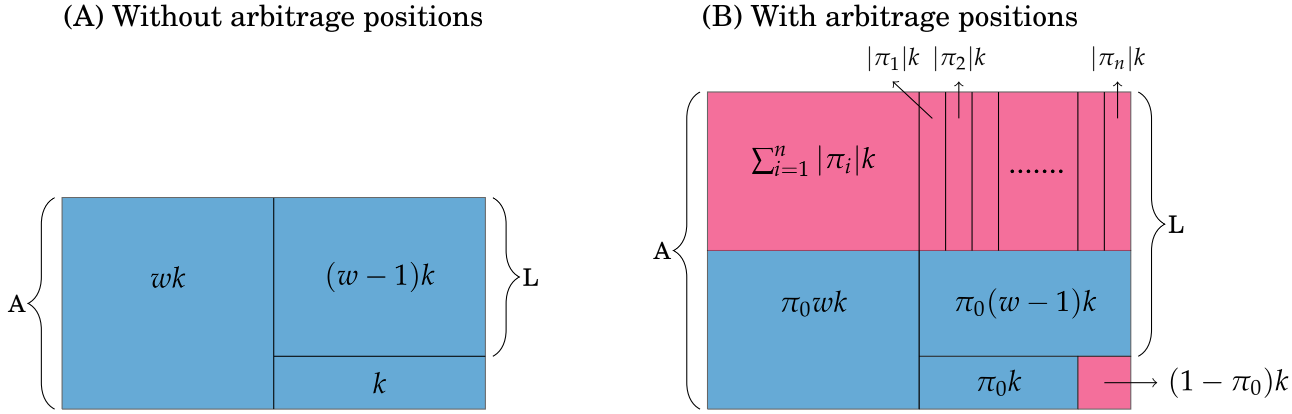


Figure 2: Arbitrageurs' balance sheet with and without arbitrage positions (A: asset, L: liability).

$\mathrm{d}\widetilde{r}_t = (\mu_t \mathrm{d}t + \sigma_t dz_t)$ where $\{z_t\}_{t=0}^{\infty}$ is a standard Brownian motion on a complete probability space. In other words, the date-t expected rate of return of this risky project is μ_t and its volatility being σ_t. In the context of currency markets, large dealer banks play an essential role in FX arbitrage. If we treat them as the arbitrageurs, this risky project represents the a consolidated portfolio of their business activities (e.g., consumer financing, commercial banking, investment banking, security brokerage and trading, asset management, etc.), in addition to FX arbitrage.

Capital accumulation without arbitrage opportunities. Ignoring the arbitrage opportunities for now, with date-t capital k_t, arbitrageurs choose their positions in the risk project and the risk-free asset. Denote by w_t the ratio of risky project investments to total capital, their investment return within the time interval $[t, t + \mathrm{d}t]$ is

$$\mathrm{d}r(w_t) = w_t(\mathrm{d}\widetilde{r}_t) + (1 - w_t)(r_t \mathrm{d}t),$$

where $\mathrm{d}\widetilde{r}_t = (\mu_t \mathrm{d}t + \sigma_t dz_t)$ by assumption. Their capital evolves according to $k_{t+\mathrm{d}t} = k_t[1 + \mathrm{d}r(w_t)] - y_t \mathrm{d}t$, that is,

$$\frac{\mathrm{d}k}{k} = r\mathrm{d}t + w(\mu - r)\mathrm{d}t + w\sigma \mathrm{d}z - \frac{y}{k}\mathrm{d}t.$$

Following the literature (e.g., He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014)), I expect $w_t > 1$, which means arbitrageurs build up leveraged positions in the risky project, funded by risk-free debt. Their balance sheet is illustrated in Panel (A) of Figure 2.

Capital accumulation with arbitrage opportunities. With arbitrage opportunities, arbitrageurs' date-t problem can be thought of as making two sets of decisions. On the one hand, they choose the amount of capital, denoted by $\pi_{0t}k_t$ ($\pi_0 \leq 1$), to support their "normal lines of business", that is, borrowing at the risk-free rate and making leveraged investment in the risky project. This investment, costing $\pi_{0t}k_t$ initially, leads to $\pi_{0t}k_t[1 + \mathrm{d}r(w_t)]$ amount of capital at date $(t + \mathrm{d}t)$, where

$\mathrm{d}r(w_t)$ follows the same definition above. The risk exposure w_t is chosen optimally under the standard risk-return trade-off. On the other hand, arbitrageurs also decide the size of arbitrage positions relative to their capital, denoted by the vector $\boldsymbol{\pi}_t = (\pi_{1t}, \ldots, \pi_{nt})^\top$, for each of the n arbitrage opportunities. Total arbitrage profits at date $(t + \mathrm{d}t)$ are $(\sum_{i=1}^{n} \pi_{it} b_{it} \mathrm{d}t)\, k_t$, or $(\boldsymbol{\pi}_t^\top \boldsymbol{b}_t \mathrm{d}t) k_t$ for simplicity, where $\boldsymbol{b}_t = (b_{1t}, \ldots, b_{nt})^\top$. These arbitrage profits come at zero cost at date-t. We can write down arbitrageurs' total capital at date $(t + \mathrm{d}t)$ as

$$k_{t+\mathrm{d}t} = k_t + \pi_{0t} k_t [1 + \mathrm{d}r(w_t)] - \pi_{0t} k_t + (\boldsymbol{\pi}_t^\top \boldsymbol{b}_t \mathrm{d}t) k_t - 0 - y_t \mathrm{d}t,$$

which extends equation (3) under the new dynamic stochastic environment with multiple arbitrage opportunities. Simplifying the equation above, arbitrageurs' capital evolves according to

$$\frac{\mathrm{d}k}{k} = \pi_0 \left[r\mathrm{d}t + w(\mu - r)\mathrm{d}t + w\sigma \mathrm{d}z \right] + \boldsymbol{\pi}^\top \boldsymbol{b} \mathrm{d}t - \frac{y}{k}\mathrm{d}t. \tag{11}$$

Panel (B) of Figure 2 illustrates the structure of arbitrageurs' balance sheet after building up arbitrage positions. Their original balance-sheet composition are colored in blue and arbitrage positions are colored in red. Taking advantage of arbitrage opportunities potentially leads to downsizing the normal business. In doing so, arbitrageurs are effectively setting a fraction $(1 - \pi_0)$ of their capital aside in storage, earning zero net returns. We can also view this amount of capital as necessary capital buffers to support arbitrage positions (thus also colored in red). In the context of major FX dealer banks, $(1 - \pi_0)k$ represents the amount of bank capital deployed to their trading desks dedicated to CIP arbitrage. The choice regarding π_0 can also be interpreted as resource allocation decisions in internal capital markets.

To sum up, arbitrageurs now choose i. π_0 that determines the size of their "conventional" investment as well as its leverage w, ii. arbitrage positions $\boldsymbol{\pi}$ in each of the arbitrage opportunities, iii. consumption rate y, to maximize their utility (10) subject to the capital accumulation equation (11).

Without arbitrage limits, arbitrageurs will choose $\pi_0 = 1$ and $|\pi_i| \to \infty$ for any $i \in \{1, \ldots, n\}$ such that $b_i \neq 0$.

Arbitrage limits arise from financial constraints defined by the set $\mathcal{C}$. Combinations of π_0 and $\boldsymbol{\pi}$ must fall within $\mathcal{C}$. Arbitrageurs face the trade-off between chasing larger arbitrage profits and downsizing their normal business operations, under this constraint. Extending equation (9), the arbitrage profit function (the support function of $\mathcal{C}$) is now defined as

$$S_{\mathcal{C}}(r, \boldsymbol{b}) = \sup_{(\pi_0, \boldsymbol{\pi}) \in \mathcal{C}} \{\pi_0 r + \boldsymbol{\pi}^\top \boldsymbol{b}\}.$$

Time-varying financial constraint. I allow for time variation in the set $\mathcal{C}$ to reflect changing financial constraints. Specifically, I assume

$$(\pi_{0t}, \boldsymbol{\pi}_t) \in \mathcal{C}_t. \tag{12}$$

Arbitrage profit functions can be defined for each $\mathcal{C}_t$ accordingly.

I collect assumptions in the previous section about the constraint set and present a summarized one below:

Assumption 5. *At any time t, $\mathcal{C}_t$ is a bounded, closed and convex subset of $[0, 1] \times \mathbb{R}^n$, which is nonempty with $(1, \mathbf{0}_n) \in \mathcal{C}_t$; the arbitrage position $\pi_{it} = 0$ when $b_{it} = 0$.*

Hedgers. I extend the hedging demand specification of (5) to each of the n arbitrage opportunities by assuming

$$\boldsymbol{q}_t = \gamma_{0,t} - \gamma \boldsymbol{b}_t, \quad \gamma > 0, \tag{13}$$

where elements in $\boldsymbol{q}_t = (q_{1t}, \ldots, q_{nt})^\top$ are external (net) hedging demands. Following the interpretations in Section 2 in the context of CIP arbitrage, q_{it} represents the demand for forward dollars via the exchange of currency i. The vector $\gamma_{0,t} = (\gamma_{01,t}, \ldots, \gamma_{0n,t})^\top$ captures the fundamental hedging demand differences among currencies, due to trade imbalances or cross-border investment. The positive scaler γ indicates that hedging demands for forward dollars are always decreasing in the CIP deviations. Appendix C further discusses micro-foundation of this specification.

The equilibrium arbitrage yield. At time t, the equilibrium arbitrage yield $\boldsymbol{b}_t^*$ is a vector such that

$$\boldsymbol{\pi}_t k_t = \boldsymbol{q}_t. \tag{14}$$

where $\boldsymbol{q}_t$ is defined by (13); $\boldsymbol{\pi}_t$ are (part of) the solutions to the arbitrageurs' problem: choosing $\{y_t, w_t, \pi_{0t}, \boldsymbol{\pi}_t\}$ to maximize the utility function (10) subject to the capital accumulation equation (11) under the constraint (12).[25]

Before characterizing equilibrium outcomes, I add three remarks to finish describing the model setup. First, the model does not consider risky arbitrage. As a result, it is not suitable for investigating many intriguing pricing phenomena such as stock market anomalies. For CIP arbitrage, instead of treating it as risky payoffs, standard practices apply valuation adjustments, modify margin requirements, or resort to VaR calculations to address risk-related concerns (e.g., the counterparty risk and the mark-to-market valuation risk). Financial constraints $\mathcal{C}_t$ in the current model encompass these scenarios, as illustrated by examples discussed in the previous section. In addition, CIP ar-

[25]Assumption A.1 in Appendix A.3 summarizes additional technical assumptions regarding investment opportunities (r_t, μ_t, σ_t), financial constraints $\mathcal{C}_t$, and external hedging demands $\gamma_{0,t}$.

bitrage does not involve convergence trading and is not subject to the (endogenous) risk induced by random arbitrage horizons in models such as Kondor (2009). Thus, this riskless arbitrage model tends to be a good fit for studying CIP deviations and other "near-arbitrage" bases, such as the positive gap between the interest on excess reserve rate and the reverse repo rate.

Second, the log-utility assumption, although inducing myopic behaviors, is not restrictive for analyzing riskless arbitrage. The intuition is that arbitrageurs cannot exploit riskless arbitrage opportunities to hedge against future shocks to their assets and financial constraints. In other words, they only adjust arbitrage positions in response to contemporaneous shocks. On the other hand, arbitrageurs do adjust risk exposures through their positions on the risky project, taking into consideration their changing investment opportunities. Appendix A.3 presents and characterizes equilibrium outcomes of the same model under general CRRA utility specifications to clarify these points.

Third, the risky project in the model prevents us from carrying over model solutions of Section 2 directly. To see this more clearly, in equation (4) of the previous section, arbitrageurs' log investment return is $\log(1 + \pi_0 r + \pi b)$. Maximizing it under the constraint $(\pi_0, \pi) \in \mathcal{C}$, it is almost trivial to see that, in equilibrium, $\pi_0 r + \pi b = S_{\mathcal{C}}(r, b)$. Under the full model here, arbitrageurs' instantaneous (expected) log investment return is $\mathbb{E}\log[1 + \pi_0 \mathrm{d}r(w) + \boldsymbol{\pi}^\top \boldsymbol{b}\mathrm{d}t]$ (ignoring consumption, long-horizon log investors effectively maximize expected log returns period by period). Maximizing it under the constraint (12) is not straightforward. I develop theoretical tools to solve this type of problems building on the concept of convex conjugacy (also see Appendix A.3 for details). These tools also apply to the "true" dynamic setting under CRRA utilities. As a preview, the results are surprisingly simple: a multivariate generalization $\pi_0 r + \boldsymbol{\pi}^\top \boldsymbol{b} = S_{\mathcal{C}}(r, \boldsymbol{b})$ still holds in equilibrium and risk exposures of arbitrageurs are adjusted through changing w (although π_0 also affects the size of their risky positions, it is completely pinned down by $\mathcal{C}$).

3.2 Equilibrium characterization

This section characterizes the equilibrium outcomes. I first present arbitrageurs' optimal choices and their capital dynamics in equilibrium. Then I show the equation that equilibrium arbitrage yields must satisfy and analyze its properties. Again, most time subscripts are suppressed.

Recall that arbitrageurs' choice variables include their consumption rate y, the amount of capital used for their "standard business" π_0, their risky asset weight w, and the vector $\boldsymbol{\pi}$ determining their arbitrage positions. I start presenting their choices of π_0 and $\boldsymbol{\pi}$ in the following proposition.

Proposition 3. *If $\mathcal{C}$ is such that $S_{\mathcal{C}}$ is differentiable, equilibrium arbitrage positions are given by*

$$\pi_i = \frac{\partial S_{\mathcal{C}}(r, \boldsymbol{b})}{\partial b_i}, \text{ for all } i = 1, \ldots, n.$$

In equilibrium, the fraction of capital maintained for investment opportunities other than arbitrage *is given by*

$$\pi_0 = \frac{\partial S_{\mathcal{C}}(r, \boldsymbol{b})}{\partial r}.$$

Optimal arbitrage positions are not affected by the profile of the risky project (i.e., its risk and return captured by μ and σ), but determined fully by the risk-free rate r in combination with arbitrage yields $\boldsymbol{b}$. The shape of $\mathcal{C}$ determines the specific functional form of $\boldsymbol{\pi}$ with regard to "prices" $(r, \boldsymbol{b})$ via the arbitrage profit function $S_{\mathcal{C}}$. Examples of this function are available in Figure 1 discussed in the previous section.

According to Proposition 3, arbitrageurs have to set aside $(1 - \partial S_{\mathcal{C}}(r, \boldsymbol{b})/\partial r)$ fraction of their total equity capital to support their optimal choice of arbitrage positions. This choice is again not affected by the risk and return characteristics defined by μ and σ. The remaining fraction will be used for building up risky asset positions of size $w(\partial S_{\mathcal{C}}(r, \boldsymbol{b})/\partial r)$.

In Appendix A.3, I show that the optimal π_0 and $\boldsymbol{\pi}$ given by Proposition 3 do not change for general CRRA utility functions. Proposition 4 below provides arbitrageurs' optimal choices of y and w. Its generalization for CRRA utility functions yields more complicated results, which are provide in Proposition A.1 of Appendix A.3.

Proposition 4. *In equilibrium, arbitrageurs' optimal rate of consumption y is such that $y = \rho k$; their position on the risky project $\pi_0 w$ equals $(\mu - r)/\sigma^2$, that is,*

$$w = \frac{\mu - r}{\sigma^2}\left(\frac{\partial S_{\mathcal{C}}(r, \boldsymbol{b})}{\partial r}\right)^{-1}.$$

Arbitrageurs' choice of total risky asset exposure $(\pi_0 w)$ exhibits the behavior of classical "Mertonian" demand (Merton, 1973).[26] The rate of consumption $y = \rho k$ is a standard result under the log utility.[27] We can always interpret their choices as a two-stage sequence. First, given the "price vector" $(r, \boldsymbol{b})$ and knowing their constraints $\mathcal{C}$, arbitrageurs nail down the size of their arbitrage positions $\boldsymbol{\pi}$ and set aside $(1 - \pi_0)$ fraction of their total capital in support of these arbitrage activities. Second, with the remaining $\pi_0 k$ amount of capital ready for use, arbitrageurs solve the standard consumption-saving problem with assets defined by the triplet (r, μ, σ).

I now present the dynamics of arbitrageurs' capital in equilibrium, which serves as the key identifying equation for quantitative analysis.

[26] Under the current log utility case, this quantity equals the myopic mean-variance efficient demand $(\mu - r)/\sigma^2$ (Proposition 4 above). In Proposition A.1 of the Appendix A.3, I extend the result for general CRRA utilities which account for both intertemporal hedging and endogenous dynamic risk aversion. Its form still complies with the "Mertonian" demands under intertemporal settings.

[27] For general CRRA utilities, the equilibrium ratio y/k varies according to the state of the economy. Proposition A.1 of the Appendix A.3 presents the general result.

Proposition 5. *In equilibrium, the arbitrageurs' capital evolves according to the following rule:*

$$\frac{\mathrm{d}k}{k} = \left[S_{\mathcal{C}}(r, \boldsymbol{b}) - \rho + \lambda^2\right] \mathrm{d}t + \lambda \mathrm{d}z, \tag{15}$$

where $\lambda = (\mu - r)/\sigma$ is the Sharpe ratio of the risky project available to arbitrageurs.

Proposition 5 allows for intuitive interpretations. To see this, plugging the two equilibrium conditions $y = \rho k$ and $\pi_0 w = \lambda/\sigma$ from Proposition 4 into the budget constraint (11), we have

$$\frac{\mathrm{d}k}{k} = \Big[\Big(\pi_0 r + \boldsymbol{\pi}^\top \boldsymbol{b}\Big) - \rho + \lambda^2 \Big] \mathrm{d}t + \lambda \mathrm{d}z.$$

Comparing the equation above with equation (15) in Proposition 5, we can see that, the equilibrium π_0 and $\boldsymbol{\pi}$ are such that

$$\pi_0 r + \boldsymbol{\pi}^\top \boldsymbol{b} = S_{\mathcal{C}}(r, \boldsymbol{b}).$$

The result indicates that when solving the infinite horizon optimization problem, arbitrageurs still behave as if they were solving the simple problem of maximizing $(\pi_0 r + \boldsymbol{\pi}^\top \boldsymbol{b})$ subject to $(\pi_0, \boldsymbol{\pi}) \in \mathcal{C}$, when choosing π_0 and $\boldsymbol{\pi}$ each period. This optimization problem is a multivariate extension of solutions to the simple model presented in Section 2 in which only one arbitrage opportunity exists.

Another way of looking at the dynamics of arbitrageurs' capital is through the view of Euler equations or, equivalently, stochastic discount factors (SDF). Let us define $\Lambda_t = e^{-\rho t}/y_t$. Then, under the log utility, optimal intertemporal choices of the arbitrageurs enforce that $\mathrm{d}\Lambda/\Lambda$ is an SDF, pricing the risky asset(s) available to them. As $y = \rho k$, $\Lambda_t = e^{-\rho t}/(\rho y_t)$, Proposition 5 indicates that

$$\frac{\mathrm{d}\Lambda}{\Lambda} = -S_{\mathcal{C}}(r, \boldsymbol{b})\mathrm{d}t - \lambda \mathrm{d}z.$$

The risk premium of the risky project $(\mu - r)\mathrm{d}t$ equals $\mathbb{E}[(-\mathrm{d}\Lambda/\Lambda)\mathrm{d}\widetilde{r}]$, the opposite of its return covariance with this specific SDF defined by the consumption (or capital) of the arbitrageurs. In other words, the consumption Euler equation holds in the model. Constraints on the arbitrage positions do not render arbitrageurs' intertemporal consumption and portfolio choice suboptimal. When $\boldsymbol{b}$ is a vector of zeros, that is, no arbitrage opportunity exists, $S_{\mathcal{C}}(r, \boldsymbol{b}) = r$.[28] The SDF takes the conventional form of $(-r\mathrm{d}t - \lambda \mathrm{d}z)$ in continuous time. Riskless arbitrage opportunities effectively serve as a "booster technology" to ramp up the risk-free rate available to the arbitrageurs.

The next proposition shows the system of equations determining the equilibrium level of arbitrage yields. It also discusses sufficient conditions for the existence and uniqueness of the equilibrium.

[28]Recall that Assumption 5 requires $\mathcal{C} \subset [0, 1] \times \mathbb{R}^n$, under which π_0 is always smaller than one.

Proposition 6. *The equilibrium arbitrage yields* $\boldsymbol{b}$ *under* (14) *solves*

$$\frac{\partial S_{\mathcal{C}}(r, \boldsymbol{b})}{\partial \boldsymbol{b}} k = \gamma_0 - \gamma \boldsymbol{b}. \tag{16}$$

If the scalar $\gamma > 0$ *and the set* $\mathcal{C}$ *satisfying Assumption 5 also guarantees that the arbitrage profit function* $S_{\mathcal{C}}$ *is twice continuously differentiable, a unique solution exists in a ball in* $\mathbb{R}^n$ *centered at zero with a radius* $\|\gamma_0\|_2/\gamma$.[29]

Proof of existence relies on verifying sufficient conditions for the Brouwer fixed-point theorem; uniqueness is a result of the implicit function theorem. Appendix A.3 contains details of the proof. The radius $\|\gamma_0\|_2/\gamma$ corresponds to the norm of arbitrage yields without arbitrageurs, that is, when $\boldsymbol{b} = \gamma_0/\gamma$. Under this scenario, the vector $\boldsymbol{b}$ adjusts such that all hedging demand imbalances equal zero. Arbitrageurs help reduce the overall equilibrium arbitrage yields in the sense that their norms become smaller than $\|\gamma_0\|_2/\gamma$. Arbitrage forces dampen the influences of the "raw hedging demands" γ_0 (hedging demands when no CIP deviations exist) on $\boldsymbol{b}$: the response of equilibrium $\boldsymbol{b}$ to changes in γ_0 is less than $1/\gamma$ whenever there are arbitrageurs taking advantage of the arbitrage opportunities induced by demand imbalances.

4 Empirics: testing model predictions and estimating the model

4.1 Data

Beyond CIP deviation measures described early on, I assemble data from several other sources. I collect the trade-weighted broad dollar index, the VIX index, Fed fund rates, treasury yields, euro implied volatilities (CBOE EuroCurrency ETF Volatility Index) from Federal Reserve Economic Data (FRED). I download the yield curve of RefCorp strips from Bloomberg (for calculating the dollar convenience yields following Longstaff (2004)).

I also collect bilateral trade data from IMF Direction of Trade Statistics; bilateral portfolio transaction data as well as cross-border bank claim data from the US Treasury International Capital (TIC) System; bilateral foreign direct investment data from the US Bureau of Economic Analysis.

I create a measure of arbitrageurs' capital in currency markets. It is motivated by the fact that these markets are predominantly dealer-intermediated. I consider 49 global dealer banks which are participants of semi-annual foreign exchange turnover surveys (FXS) by local monetary authorities in New York, London, Tokyo, Toronto, Sydney, Singapore, and Hong Kong. Table A1 of Appendix

[29] Of note, $\mathcal{C}$, γ_0, k and r all vary across time in the model. The equilibrium condition holds one by one at each time point. The existence and uniqueness results thus apply only to each time period for a given collection of $\{\mathcal{C}, \gamma_0, k, r\}$. The proposition is silent on the Markovian equilibrium under which we are interested in the property of a mapping from the state of the economy to equilibrium arbitrage yeilds $\boldsymbol{b}$ such that the equation in this proposition always holds. I leave this exploration for future research.

D lists names of their holding companies. The equity capital of these dealer banks' holding companies is my intended measure of arbitrageurs' capital. Their fundamental and price data come from Compustat and CRSP. I use Bloomberg to access their five-year credit default swap (CDS) rates.

4.2 Supporting evidence of the model

Without committing to specific financial constraints, the model still yields a strong prediction: an increasing arbitrage yield (e.g., the CIP deviation) should predict higher returns on arbitrageurs' capital, and as the former goes up, the latter should go up increasingly fast (a convex relationship). The equilibrium outcome stated in equation (15) illustrates this point. On the left-hand side of this equation are arbitrageurs' capital returns next period, and on the right-hand, the function $S_{\mathcal{C}}$, which is increasing and convex in $\boldsymbol{b}$, the arbitrage yields. It is worthwhile reiterating that the convexity of this function is a direct result of Assumption 6 that $\mathcal{C}$ is always convex.

4.2.1 Arbitrageurs' capital returns in currency markets

The most direct measure of the 49 FX dealer banks' equity capital is the book equity (BE) of their holding companies. As suggested by the theory, if these banks are indeed *the* arbitrageurs of FX derivatives markets, CIP deviations should predict returns on their equity capital. I compute for each bank their book equity returns (growth of book equities next quarter divided by present book equity levels) and estimate the following panel regressions:

$$\frac{1}{\tau}\mathrm{return}_{i,t+\tau} = \alpha_i + \beta\,\bar{b}_t + \epsilon_{i,t+\tau},$$

where returns on the left hand side are annualized by dividing $\tau = 0.25$ (a quarter), the subscript i denotes banks and t stands for quarters. The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF (namely, the G6 currencies) against the dollar. Sample periods begin from March 2009 and end at December 2019.[30] The first two columns in Table 2 show the regression results. Overall, average CIP deviations significantly predict these banks' book equity growth: one standard deviation increase in the deviations is associated with around 1.6 percentage points increase in FX dealer banks' book equity. This finding is robust to measurements of CIP deviations using either the currency swap rates or the forward-OIS implied basis.

There is a major drawback in using the book equity measure: it is an accounting variable that is observable only quarterly. This drawback will become more pronounced as later model testing and estimation include nonparametric procedures. To circumvent this issue, a the potential surrogate measure, the market equity (ME), becomes particularly attractive. This measure comes from high

[30] I use this sample period to avoid tumultuous periods of the global financial crisis and the COVID-19 pandemic.

quality real-time market price data. It is also worthwhile noting that, for the 49 FX dealer banks under study, their average and median market-to-book (MB) ratios equal 1.10 and 1.05 respectively (during the sample period of 2009-2019).[31] The time-series standard deviation of market-to-book ratios averaged across these banks is 0.13. These features partially motivate the use of market equity.

An additional motivation for using the market equity measure (at least in the context of testing the predictive relationship here) is the fact that CIP deviations *do not* predict changes in the market-to-book ratios. As the equation

$$\frac{\mathrm{BE}_{t+1}}{\mathrm{BE}_t} \times \frac{\mathrm{MB}_{t+1}}{\mathrm{MB}_t} = \frac{\mathrm{ME}_{t+1}}{\mathrm{ME}_t}$$

holds by definition, if a predictor does not predict the ratio $\mathrm{MB}_{t+1}/\mathrm{MB}_t$ which stands for "returns" on the book-to-market ratio, it must simultaneously predict (or fail to predict) book and market equity returns. The third and fourth columns of Table 2 verify this conclusion by regressing $(\mathrm{MB}_{t+1}/\mathrm{MB}_t - 1)$ on $\bar{b}_t$. The slope coefficients are statistically indistinguishable from zero. Since we have already seen from the same table that $\bar{b}_t$ predict book equity returns of the 49 FX dealers, this variable should also predict their market equity returns.

Now I redo the panel regression using market equity returns. Both measures of CIP deviations are considered. For comparison, I still consider quarterly observations of quarterly returns. The last two columns of Table 2 document the results. Average CIP deviations also significantly predict these banks' market equity returns. The slope coefficients are larger: one standard deviation increase in the deviations is associated with six percentage points increase in expected market equity returns. The larger regression coefficient (compared with the case for book equity returns) is mainly due to the fact that market equity returns are more volatile than book equity returns: annualized time-series volatilities are 28.8% for the former and 9.8% for the later.

From now on, I will use returns on market equity of the 49 FX dealer banks' holding companies to measure arbitrageurs' capital returns in the model. Using market returns raises concerns about confounding effects of other return predictors, such as valuations ratios and volatilities. In the following section, I will further investigate the predictive relationship and try to mitigate these concerns by adjusting for potential return predictors.

4.2.2 CIP deviations predict arbitrageurs' capital returns

I document additional evidence on the predictive power of CIP deviations on FX dealer banks' capital returns. Columns headed with "FXS" in Table 3 present results from the following time-series

[31] Market-to-book ratios of bank equities are around one not only during the post crisis period under study. In fact, these ratios have been close to one until the mid 1990s. During the exceptional period of 1996-2008, MB ratios of banks were over two. Explanations to these patterns are beyond the scope of this paper. Interested readers may refer to papers such as Calomiris and Nissim (2014); Atkeson et al. (2019).

regressions using quarterly observations:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta\,\overline{b}_t + \epsilon_{t+\tau},$$

where the dependent variable is one-quarter-ahead ($\tau = 0.25$) value- or equal-weighted stock returns of the 49 dealer banks' holding companies. Returns are annualized by dividing τ. The independent variable $\overline{b}_t$ is still the cross-sectional average of absolute one-year basis swap rates of G6 currencies against the dollar. Sample periods begin from March 2009 and end at December 2019. Regression coefficients are statistically significant in these columns of Table 3. On average, one basis point increase in the average CIP deviations predicts around two percentage points increase in the returns of arbitrageurs' capital.

A set of placebo tests are included Table 3. The same predictive regressions are repeated for returns of five exchange-traded funds (ETFs) tracking the S&P500 index (SPY), the global financial sector (IXG), the US financial sector (IYF), US broker-dealers and securities exchanges (IAI), and US insurance companies (KIE). Average CIP deviations *do not* predict placebo outcomes, except for returns of the ETF tracking the global financial sector. This unique positive finding is not surprising as the 49 FX dealer banks are likely to be essential constituents of the fund. These placebo tests suggest that the 49 global dealer banks under consideration do play special roles in CIP arbitrage: they tend to be *the* arbitrageurs both in my model and in reality.

Table 4 assembles additional results for the same time-series regression using daily and monthly observations. Regression coefficients are remarkably stable for the main outcome variable: value-weighted equity returns on the 49 FX dealers banks' holding companies. One basis point increase in the average CIP deviations is still associated with around two percentage point increase in these returns. Placebo test results remain consistently negative (again, except for the ETF tracking the global financial sector). For monthly observations, five hedge fund index returns are also included for placebo tests: one global composite index from BarclaysHedge, four indices from Hedge Fund Research tracking global composite, relative value arbitrage, global-macro, and macro-currency strategies. CIP deviations do not predict the composite hedge fund return indices.[32]

Table 5 presents results from the adjusted version of the predictive regression:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta\,\overline{b}_t + \phi \cdot \text{control}_t + \epsilon_{t+\tau},$$

in which control variables include earnings yields and dividend yields averaged across the 49 FX dealer banks' holding companies. Quarterly returns are annualized by dividing $\tau = 0.25$. The effective fed fund rates, as well as the VIX index are also incorporated. Results for both monthly and

[32]All results till now focus on quarterly returns. Table A2 in Appendix D also confirms the predictive relationship (as well as negative results from placebo tests) for *monthly* returns. The regression coefficients remain stable (around two) for the main outcome variable, though adjusted R^2s drop for monthly returns.

daily observations are reported. Average G6 currency CIP deviations still demonstrate significant predictive power, but the magnitude is reduced by a half after controlling for the banks' earnings yields, which also strongly predict the returns. Table A3 reports results from repeating the same exercise for monthly returns. All results remain largely unchanged, except for the declined adjusted R^2s. To sum up, this set of time-series regressions suggest that CIP deviations predict arbitrageurs' capital returns, both before and after controlling for common return predictors. In addition, banks' earnings yields also emerge as an important return predictor.

4.2.3 The predictive relationship is convex

I now further investigate whether the predictive relationship is convex, as suggested by equation (15) in Proposition (5). As the term $S_{\mathcal{C}}(r, \boldsymbol{b})$ contains both the risk-free rate and the CIP deviations, I rewrite equation (15) as follows:

$$\frac{1}{\mathrm{d}t}\frac{\mathrm{d}k/k}{r} = S_{\mathcal{C}}\left(1, \frac{\boldsymbol{b}}{r}\right) - \frac{\rho}{r} + \frac{\lambda^2}{r} + \frac{\lambda \mathrm{d}z}{\mathrm{d}t},$$

after dividing both sides by $r\mathrm{d}t$ and leveraging the property of $S_{\mathcal{C}}$ that it is positively homogeneous of degree one. This motivates the following regression specification

$$\frac{1}{\tau}\left(\frac{\text{return}_{t+\tau}}{r_t}\right) = S_0\left(\frac{\bar{b}_t}{r_t}\right) + \phi \cdot \text{control}_t + \varepsilon_{t+\tau},$$

in which r_t denotes the risk-free rate; $\tau = 0.25$ denotes the time interval of one quarter; controls include the reciprocal of r_t (as suggested by the theory in which $-\rho/r$ shows up), the earnings yield that emerges as a strong return predictor in the previous section, as well as the VIX index; the function $S_0(\cdot)$ captures the functional form of $S_{\mathcal{C}}(1, \cdot)$.

To begin with, I contrast the parametric configuration of $S_0(x) = \beta_0 + \beta x$ and $S_0(x) = \beta_0 + \beta x^2$ in Table 6.[33] The slope coefficient β is significantly positive *only* under the quadratic specification. Estimates of these coefficients are stable across daily and monthly observations. To mitigate concerns about low risk-free rates creating large dependent variables to the extent that some may become "outliers", I redo the same regressions using robust estimators based on the Huber loss function. Robust estimators confirm that β is only significant under the quadratic specification, suggesting a convex predictive relationship.

Next, I estimate the equation using semi-parametric techniques. The nonparametric component S_0 is expanded to shape-constrained B-spline basis (Eilers and Marx, 1996). Table 7 reports the estimation results and tests for the significance of $S_0(\bar{b}_t/r_t)$ using both daily and monthly observations. I

[33] I do not incorporate the linear and quadratic terms simultaneously due to the potential multicollinearity concern: correlation between $\bar{b}/r$ and $(\bar{b}/r)^2$ is 0.9 in the sample.

consider three types of configurations for S_0: i. both convex and increasing (as suggested by theory), ii. only increasing, and iii. no restrictions. Under all conditions, regression coefficients for $1/r_t$ are significantly negative, as suggest by the theory. The nonparametric term $S_0(\bar{b}_t/r_t)$ cannot be ignored. Figure 5 plots the estimated $S_0(x)$ (which equals $S_{\mathcal{C}}(1, x)$) under the three specifications. Convex patterns consistently show up even without imposing the convexity constraint. A kink exists around $x = 3$, before which S_0 increases slowly and after which the function shoots up, indicating substantial arbitrage profits when $|b| > 3r$. As the median level of r is 16 basis points, arbitrageurs appear to enjoy large arbitrage profits after CIP deviations exceed 48 basis points.

This semi-parametric estimation implicitly adopts one crucial assumption: the financial constraint $\mathcal{C}$ does not change across time. Next, I will account for the dynamics of financial constraints and formally estimate the model.

4.3 Quantitative specifications, identification, and estimation

In this section I enrich the theoretical model presented above with additional assumptions to map it to data. The main goal is to quantify $\mathcal{C}_t$ (financial constraints) and $(\gamma_{0,t}, \gamma)$ (hedging demands and the elasticity parameter) in equation (16) of Proposition 6. To achieve this goal, I adopt a two-step estimation strategy. First, I estimate $\mathcal{C}_t$ using the equilibrium capital accumulation equation (15). Then, knowing $\mathcal{C}_t$ and thus the function $S_{\mathcal{C}_t}$, I compute the equilibrium arbitrage positions on the left hand side of equation (16), and then estimate hedging demands. I begin with an assumption simplifying the financial constraints.

4.3.1 Separating shapes and dynamics of financial constraints

Time-varying financial constraints $\{\mathcal{C}_t\}_{t\geq 0}$ is a series of sets satisfying Assumption 5. Estimating a sequence of random sets is challenging (if not impossible). To make progress, I adopt a simplifying assumption about the financial constraints by separating their shapes and dynamics.

Assumption 6. *There exists a constant set $\mathcal{C}_0$, such that $\mathcal{C}_t = \{(\pi_0, \alpha_t \boldsymbol{\pi}) : (\pi_0, \boldsymbol{\pi}) \in \mathcal{C}_0\}$ for a sequence $\alpha_t > 0$.*

This assumption implies that at any time, financial constraints defined by the set $\mathcal{C}_t$ is derived from a "baseline" $\mathcal{C}_0$ by shifting the largest possible arbitrage positions. The time series $\{\alpha_t\}$ serves the role of "shifters", which captures variation in the financial constraints. When $\alpha_t > 1$, $\mathcal{C}_t$ subsumes the baseline specification $\mathcal{C}_0$, larger arbitrage positions become feasible conditional on the same amount of capital dedicated to arbitrage activities (π_0 fixed). When $0 < \alpha_t < 1$, $\mathcal{C}_t$ shrinks, and arbitrageurs tend to cut back their arbitrage positions.

Under Assumption 6, the shape and dynamics of financial constraints each has a concrete characterization: the set $\mathcal{C}_0$ for the shape and the sequence $\{\alpha_t\}$ for the dynamics. I discuss this dissection

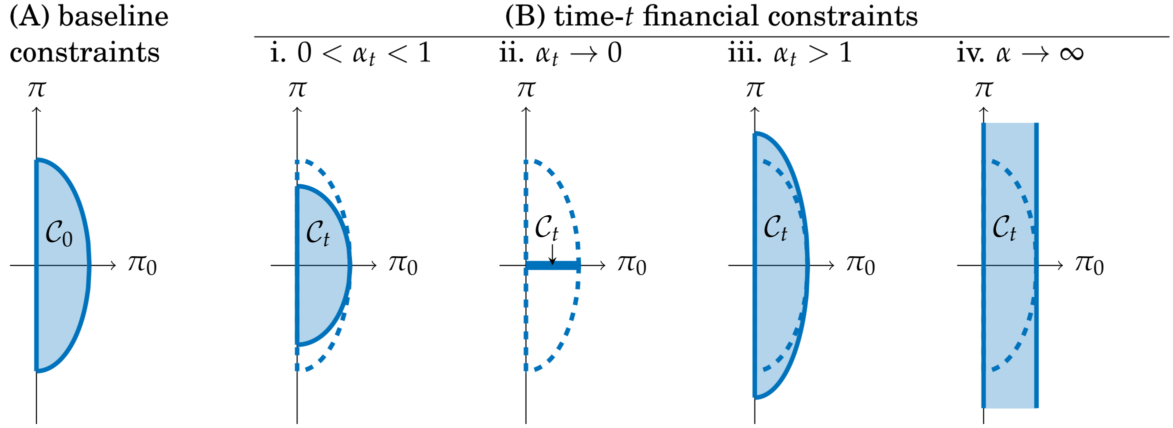


Figure 3: Dissecting the shape and dynamics of financial constraints

of financial constraints intuitively through Figure 3.[34] The first plot to the left of Figure 3 shows the baseline constraints defined by the set $\mathcal{C}_0$. This set determines the shape of financial constraints. In this illustration, it represents a VaR condition applied to the arbitrage positions (see examples from the previous section for details). $\mathcal{C}_0$ can also depict other types of constraints or combinations of multiple constraints. The other four plots in Figure 3 describe how a sequence of sets $\{\mathcal{C}_t\}$ is generated from combining $\mathcal{C}_0$ and $\{\alpha_t\}$. The time series $\{\alpha_t\}$ translates to the dynamics of financial constraints, according to Assumption 6. Plot (B)-i. and (B)-iii of Figure 3 illustrate how the financial constraints become tighter or loser from their baseline level according to the value α_t ($\mathcal{C}_0$ boundaries outlined in the dashed curves for comparison). Plot (B)-ii. and (B)-iv of Figure 3 are two extreme cases. Under the first scenario, α_t goes to zero and the set $\mathcal{C}_t$ collapses to a line segment: no arbitrage activities are allowed. All hedging demand imbalances have to be counterbalanced by large arbitrage yields. Under the second scenario, α_t becomes infinitely large, and the constraints morph into a band spanning to infinity: no limits to arbitrage exist. This corresponds to the frictionless benchmark, under which CIP deviations must always be zero.

The arbitrage profit function for time-t financial constraints $\mathcal{C}_t$ under Assumption 6 is given by the following lemma.

Lemma 1. *Under Assumption 6,* $S_{\mathcal{C}_t}(r, \boldsymbol{b}) = S_{\mathcal{C}_0}(r, \alpha_t \boldsymbol{b})$.

Lemma 1 translates Assumption 6 on sets $\{\mathcal{C}_t\}$ into properties of the arbitrage profit function. The baseline set $\mathcal{C}_0$ determines the functional form of $S_{\mathcal{C}_0}$ (shape); the series $\{\alpha_t\}$ induce time variation to the financial constraints, as well as arbitrage profit functions (dynamics). Quantifying the financial constraints is equivalent to estimating both the function $S_{\mathcal{C}_0}$ and the sequence α_t.

[34]For the ease of exposition, the illustrations cover the case of one arbitrage opportunity, while the intuitions easily carry over to higher dimensions.

4.3.2 Parameterization

I introduce a simplifying assumption and parameterize three components of the model to facilitate estimation, summarized by four items in this section.

Item 1: reducing the dimension of financial constraints. I begin by simplifying the (baseline) shape of financial constraints, defined via the function $S_{\mathcal{C}_0}$. The challenge to estimate this object comes from the "curse of dimensionality": as a rule of thumb, estimating a function of dimension d generally requires a sample size that is an exponential of d (Stone, 1982). I adopt the following simplifying assumption to sidestep this challenge.

Assumption 7. *$S_{\mathcal{C}_0}(1, \boldsymbol{b}) = S_0(\bar{b})$ where $\bar{b} = \sum_{i=1}^{n} w_i |b_i|$ and $\sum_{i=1}^{n} w_i = 1$.*

Interpretation of the assumption is straightforward. It treats the weighted average of CIP deviations as a measure of overall arbitrage yields accessible to arbitrageurs. I use over-the-counter FX derivatives trading volume to construct these weights. The derivatives include FX forwards, FX swaps, and currency swaps. The trading volume data also come from semi-annual FX surveys of local monetary authorities in New York, London, Tokyo, Toronto, Sydney, Singapore, and Hong Kong. Figure A1 in the Appendix plot the volume shares of G6 currencies and the remainder, beginning from the year 2009. Though I suppress time subscripts here, these weights can vary across time when used for aggregating CIP deviations at different time (at time t, $\bar{b}_t = \sum_{i=1}^{n} w_{it} |b_{it}|$).

Figure 4 demonstrates implications from Assumption 7 in detail. The function $S_0(x)$ defines a support function

$$S_{\mathcal{C}_0^{2D}}(x, y) = x S_0(y/x)$$

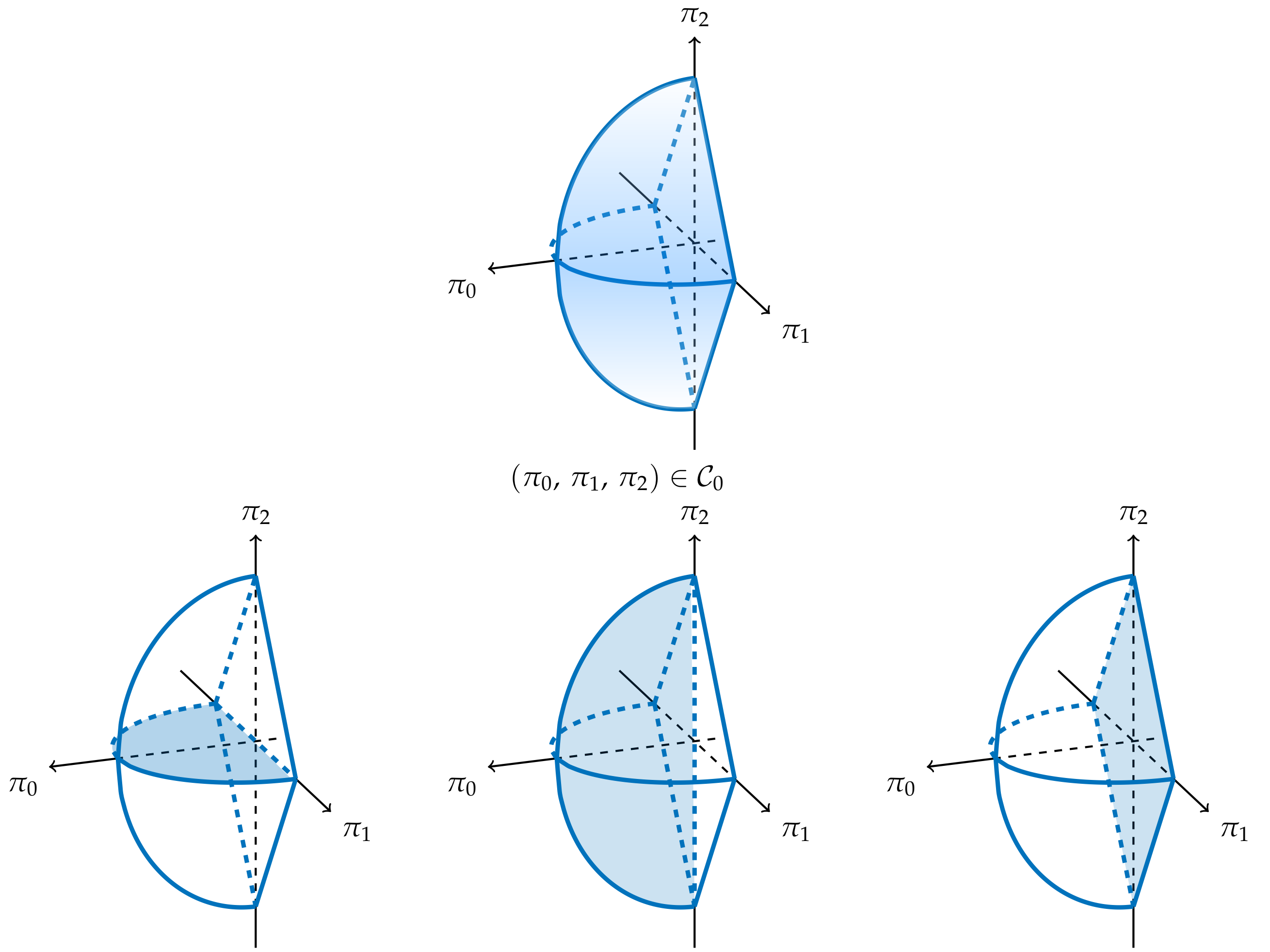
for a set $\mathcal{C}_0^{2D}$ in $\mathbb{R}^2$. Combing this two-dimensional set with a vector of weights further generates the financial constraint $\mathcal{C}_0$ in $\mathbb{R}^{n+1}$. The top plot in Figure 4 illustrates a three-dimensional set $\mathcal{C}_0$, the configuration of which satisfies Assumption 7. The three plots at the bottom show its intersections with three planes $\{(\pi_0, \pi_1, \pi_2) : \pi_i = 0\}$, $(i = 0, 1, 2)$. In the π_0-π_1 (sub)space, the intersection is indeed a set defined as $\{(\pi_0, w_1\pi_1) : (\pi_0, \pi_1) \in \mathcal{C}_0^{2D}\}$. The same rule holds for the intersection in the π_0-π_2 (sub)space. The weighted average term in Assumption 7 is reflected directly in the π_1-π_2 intersection (see the diamond shape in the plot at the bottom right corner).

Combining Assumption 7 with Lemma 1,

$$S_{\mathcal{C}_t}(r_t, \boldsymbol{b}_t) = S_{\mathcal{C}_0}(r_t, \alpha_t \boldsymbol{b}_t) = r_t S_{\mathcal{C}_0}\left(1, \frac{\alpha_t \boldsymbol{b}_t}{r_t}\right) = r_t S_0\left(\alpha_t \frac{\bar{b}_t}{r_t}\right). \tag{17}$$

Substituting this result into equation (15), and dividing both sides by $r_t \mathrm{d}t$, we have

$$\frac{1}{\mathrm{d}t} \frac{\mathrm{d}k_t / k_t}{r_t} = \left[S_0\left(\alpha_t \frac{\bar{b}_t}{r_t}\right) - \frac{\rho}{r_t} + \frac{\lambda_t^2}{r_t} \right] + \frac{\lambda_t}{r_t} \frac{\mathrm{d}z_t}{\mathrm{d}t}. \tag{18}$$



$(\pi_0,\, \pi_1,\, \pi_2) \in \mathcal{C}_0$

$\{(\pi_0,\, w_1\pi_1) : (\pi_0,\, \pi_1) \in \mathcal{C}_0^{2D}\}$ $\{(\pi_0,\, w_2\pi_2) : (\pi_0,\, \pi_2) \in \mathcal{C}_0^{2D}\}$ $w_1|\pi_1| + w_2|\pi_2| \leq \sup_{(0,\pi)\in\mathcal{C}_0^{2D}} \pi$

Figure 4: The (baseline) financial constraint $\mathcal{C}_0$ and its two-dimensional generator $\mathcal{C}_0^{2D}$ under Assumption 7.

I now introduce two additional parameterization schemes for objects in equation (18): the Shape-ratios of arbitrageurs' risky project λ_t and the dynamics of financial constraints α_t.

Item 2: parameterizing Shape-ratios. I parametrize the whole term $(\lambda_t^2/r_t - \rho/r_t)$ as the linear combination of variables that may predict the arbitrageurs' capital return $\mathrm{d}k_t/k_t$ other than the CIP deviations, that is, $\lambda_t^2/r_t - \rho/r_t = \boldsymbol{\phi}^\top \boldsymbol{v}_t$. The vector $\boldsymbol{v}_t$ include variables such as earnings yields for the 49 dealer banks, and the VIX index, which are potential predictors of the capital return $\mathrm{d}k_t/k_t$. The reciprocal of r_t is also included as suggested by theory.

Item 3: parameterizing the dynamics of financial constraints. I parameterize the positive process α_t as $\exp(\boldsymbol{\delta}^\top \boldsymbol{u}_t)$ where $\boldsymbol{u}_t$ is a vector containing variables that may drive the time-series variation in financial constraints. It includes the dollar index, quarterly lagged volatilities of average CIP deviations, changes in dealer banks' CDS, the TED spread (three-month dollar LIBOR rates minus the three-month Treasury bill rates), the implied volatility of euro, the VIX index, and the dollar convenience yield (the three-month RefCorp bond yield minus the three-month treasury yield).[35]

Under these parameterization schemes, we can now write equation (18) as

$$\frac{1}{\tau}\left(\frac{\text{return}_{t+\tau}}{r_t}\right) = \left[S_0\left(\exp(\delta^\top \boldsymbol{u}_t)\frac{\overline{b}_t}{r_t}\right) + \boldsymbol{\phi}^\top \boldsymbol{v}_t\right] + \varepsilon_{t+\tau}, \tag{19}$$

after replacing dt by τ, in which the error term $\varepsilon_{t+\tau}$ are future shocks to arbitrageurs' capital. Estimating equation (19) yields the function $S_0(\cdot)$ as well as vectors δ and $\boldsymbol{\phi}$. According to equation (17), knowledge regarding the function S_0 and the vector δ (which translates into α_t) fully reveals $S_{\mathcal{C}_t}$ (the arbitrage profit function determined by time-varying financial constraints). In equilibrium, arbitrage positions can be calculated as

$$\pi_{it} = \frac{\partial S_{\mathcal{C}_t}(r_t, \boldsymbol{b}_t)}{\partial b_{it}} = \frac{r_t \partial S_0(\alpha_t \overline{b}_t / r_t)}{\partial b_{it}} = \alpha_t w_{it} \text{sgn}(b_{it}) S_0'\left(\frac{\alpha_t \overline{b}_t}{r_t}\right), \quad \alpha_t = \exp(\delta^\top \boldsymbol{u}_t). \tag{20}$$

Now shifting attention to equation (16) and writing it in an element-wise manner, we have

$$\pi_{it} k_t = \gamma_{0,it} - \gamma b_{it}.$$

The left-hand side of this equation becomes observable if we know S_0 and δ, according to equation (20). I now introduce parameterization for hedging demand intercepts $\gamma_{0,it}$ on the right-hand side.

Item 4: parameterizing hedging demands. Hedgers' demands are further specified as follows (optimization foundation for the hedging demands in Appendix C helps motivate the specification):

$$\gamma_{0,it} = \boldsymbol{\beta}_i^\top \boldsymbol{x}_{it} + \ell_{it},$$

where $\boldsymbol{x}_{it}$ is a vector of observable hedging demand drivers including bilateral net exports, net foreign direct investment flows, net security purchases (long-term bonds and equities), changes in net cross-border bank claims, and interest rate differentials (all calculated as domestic, the US, minus foreign, country i); an intercept term of constant one is also included in $\boldsymbol{x}_{it}$; $\ell_{it} \sim \mathcal{N}(0, \sigma_\ell^2)$ captures unobservable components of hedging demands (or, in extension, liquidity-driven demands for forward dollars which I do not model explicitly in the micro-foundation section of Appendix C). This specification

[35] The dollar index captures risk-bearing capacity of global banks as argued by Avdjiev, Du, Koch, and Shin (2019). Past volatilities of CIP deviations may affect VaR calculations involving FX derivatives positions. Bank CDS rates determine funding value adjustments as illustrate by Andersen, Duffie, and Song (2019). The TED spread measures credit risk in the banking sector. I add the implied volatility of euro and the VIX index as additional controls for risk appetite in currency markets and, more broadly, global financial markets. The measurement of dollar convenience yields follows Longstaff (2004); Augustin et al. (2020) find that swap dealers' effective funding rates are related to convenience yields.

implies that

$$\begin{aligned}\pi_{it}k_t &= \boldsymbol{\beta}_i^\top \boldsymbol{x}_{it} - \gamma b_{it} + \ell_{it} \\ &= \overline{\boldsymbol{\beta}}^\top \boldsymbol{x}_{it} + \sum_{j=1}^{n-1} \boldsymbol{\eta}_j^\top \left(I[j=i] \times \boldsymbol{x}_{jt} \right) - \gamma b_{it} + \ell_{it}. \end{aligned} \tag{21}$$

The second equation in (21) adopts the transformation $\boldsymbol{\beta}_i = \overline{\boldsymbol{\beta}} + \boldsymbol{\eta}_i$ where $\sum_{i=1}^n \boldsymbol{\eta}_i = 0$. As a result, the vector $\overline{\boldsymbol{\beta}}$ is the cross-sectional average of $\boldsymbol{\beta}_i$, which accounts for mean responses of hedging demands to observables in the model for all currencies in the sample. Under this specification, estimating $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ is equivalent to estimating $\overline{\boldsymbol{\beta}}, \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_{n-1}$.

4.3.3 Identification and estimation

Under the current parameterization scheme, model estimation takes two steps. First, I estimate equation (19) to find the triplet $\{S_0(\cdot), \boldsymbol{\delta}, \boldsymbol{\phi})\}$; these estimates allow me to calculate arbitrage positions π_{it} according to equation (20). Second, knowing π_{it} as well as k_t, I estimate vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ and the "semi-elasticity" parameter γ from equation (21).[36]

Step 1: estimating financial constraints. The first step relies on the following identification assumption

$$\mathbb{E}\left[\varepsilon_{t+\tau} \mid \boldsymbol{b}_t, \boldsymbol{u}_t, \boldsymbol{v}_t\right] = 0,$$

in equation (19), which holds according to the theory. Specifically, this condition argues that the current CIP deviations ($\boldsymbol{b}_t$), dynamics of financial constraints (determined by $\boldsymbol{u}_t$), as well as drivers of arbitrageurs' expected capital returns ($\boldsymbol{v}_t$) do not affect shocks to arbitrageurs' *future* realized capital returns. This argument *does not* preclude the possibility that current (or even past) shocks to arbitrageurs' capital affect these variables. Arbitrageurs in the model do respond to contemporaneous shocks and adjust their arbitrage positions. Moreover, since they are global dealer banks, these shocks can even have "real" impacts through trade finance (Xu, 2020) and cross-border capital flows (Amiti, McGuire, and Weinstein, 2019), thus affecting the hedging demands. This two-sided influence complicates equilibrium CIP deviations and can induce co-movement between the arbitrage yields and contemporaneous shocks to arbitrageurs' capital. However, these relationships should not apply to future unexpected shocks, as the identification condition commands.

The identifying condition can be violated if, for example, additional unobservable risk premium drivers exist. To be more concrete, this corresponds to the case that the $\boldsymbol{\phi}^\top \boldsymbol{u}_t$ term in equation (19) should in fact be $(\boldsymbol{\phi}^\top \boldsymbol{u}_t + \ell_t^u)$ where ℓ_t^u is the unobservable component. This term must be correlated

[36] I call the parameter γ "semi-elasticity" because b is related to logarithms of forward prices F according to equation 1, the initial definition of CIP deviations. In addition, hedging demands in the model can be interpreted as forward dollar demands as discussed in Section 2. Of course, $(-\gamma)$ should be the proper semi-elasticity.

with $\boldsymbol{b}_t$ through its impact on k_t (recall that $\boldsymbol{b}_t$ must solve the equilibrium condition (16) at time t). Given the fact that powerful return predictors are usually difficult to find beyond valuation ratios and volatility measures (which I have included in the vector v_t), this concern might not be of primary importance.

The current framework is in fact flexible enough to incorporate additional controls that drive arbitrageurs' equity returns. Future research could help improve the current estimation when new dealer bank equity return predictors are identified, which will be added into the vector v_t.

If a meaningful unobserved risk premium driver does exist, my estimation may exaggerate the response of arbitrageurs' capital returns to arbitrage yields. This is because higher ℓ_t^u is equivalent to higher expected capital returns, and is associated with lower current capital valuations. According to result [ii] of Proposition 2, this leads to higher (absolute) CIP deviations. In other words, $\mathrm{Cov}(\ell_t^u, \bar{b}_t) > 0$. As a result, the estimated response of returns to CIP deviations will subsume both the direct effects from arbitrage profits and (positively) confounded effects through ℓ_t^u. The escalated level of arbitrage profit functions map to a more lenient view of financial constraints: all else equal, relaxed financial constraints allows arbitrageurs to build more aggressive arbitrage positions and reap larger arbitrage profits. Under such scenario, I interpret my estimates of the financial constraints as conservative ones (i.e., supersets) that must contain the truth at each time period.

With the identifying condition $\mathbb{E}\left[\varepsilon_{t+\tau} \mid \boldsymbol{b}_t, \boldsymbol{u}_t, \boldsymbol{v}_t\right] = 0$, I estimate equation (19) using semi-parametric nonlinear least squares. The algorithm for estimating this equation is described in Appendix B.

Table 8 reports estimation results from the first step. According to Table 8, increases in dollar index, lagged currency swap volatility, and implied volatility of euro are significantly associated with tightening financial constraints. I also consider equal weighting ($\bar{b}_t = \sum_{i=1}^n |b_{it}|/n$) for robustness and results remain largely unchanged.

I plot in Figure 6 the times series of α_t (adjusted by sample mean) based on the estimates of δ in Table 8 under the volume-weighting scheme. Smaller α_t indicates tighter financial constraints. According to Figure 6, arbitrageurs appear to face toughest constraints during the year 2015-2016. The sharp tightening begins from the middle of 2014. Perhaps not coincidentally, the Volcker rule regulating proprietary trading becomes in effect during the second quarter of 2014. In addition, the supplementary leverage ratio requirement is finalized during the third quarter of this year. The estimated dynamics of financial constraints seems to delineate FX Dealer banks' responses to these regulatory reforms. Another interesting period is the first quarter of 2017, witnessing extremely tight constraints. It is during the same quarter that the liquidity coverage ratio (LCR) requirement reaches its full effects.

Step 2: estimating hedging demands. In the second step, I estimate hedging demand parameters. I calculate π_{it} using equation (20) based on estimates of $S_0(\cdot)$ and δ from the first step. Now the goal is to estimate parameters $\boldsymbol{\beta}_i$ and γ in equation (21). Since the left-hand side of this equation are now observable, a standard panel-data linear regression can generate estimators for $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n, \gamma)$. The main issue with this estimation is that unobservable hedging demands ℓ_{it} will affect the equilibrium deviations b_{it}, thus contaminating the ordinary least-square estimator of γ.

To address this issue, I propose an instrumental variable (IV) for b_{it} based on the following assumption:

$$\mathbb{E}\left[\ell_{it} \mid \boldsymbol{x}_{i't}\right] = 0, \quad i' \neq i.$$

This condition states that unobservable hedging demands for a particular currency are not related to observable hedging demand drivers of *other* currencies. In other words, bilateral trade and portfolio flows between the UK and US, which may drive hedging demands for pounds, should not affect hedging demands for yen. If this condition is satisfied, we can instrument b_{it} using estimators $\widehat{b}_{it}$ from the following (first-stage) regression:

$$b_{it} = \boldsymbol{\psi}^\top \boldsymbol{z}_{it} + \overline{\boldsymbol{\phi}}^\top \boldsymbol{x}_{it} + \sum_{j=1}^{n-1} \boldsymbol{\xi}_j^\top \left(I[j=i] \times \boldsymbol{x}_{jt}\right) + e_{it},$$

because the right-hand side instrumental vector $\boldsymbol{z}_{it} = \sum_{i' \neq i} w_{i't} \boldsymbol{x}_{i't}^{(-\iota)}$ is not associated with ℓ_{it}. The weights are calculated from the volume of FX derivatives, which also appear in Assumption 7 and equation (20). The superscript "$(-\iota)$" for x means that the constant one for intercepts is excluded from this vector.

This instrument should not be a weak one in theory ($\boldsymbol{\psi} \neq \mathbf{0}$), as it directly affects levels of CIP deviations for other currencies (i.e., the vector $\boldsymbol{b}_{-i}$). Changes to $\boldsymbol{b}_{-i}$ will affect arbitrageurs' equilibrium arbitrage positions not only for the involved currencies ($\boldsymbol{\pi}_{-i}$), but also for currency i (π_i). If we conceptualize arbitrageurs as "suppliers" of arbitrage services, this instrument is effectively a supply shifter in the tradition of Berry, Levinsohn, and Pakes (1995). The volume-weights reflect the belief that demands for derivative contacts on dominant currencies should have larger impacts on arbitrageurs' optimal positions, transmitting more pronounced "supply" shocks.

The exclusion restriction of the proposed instruments can be invalid when there are common shocks to both observable hedging demand drivers $\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{nt}$, and latent hedging demands $\ell_{1t}, \ldots, \ell_{nt}$, thus relating $\boldsymbol{x}_{i't}$ to ℓ_{it}. A necessary outcome of this scenario is that $\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{nt}$ present a strong factor structure. In the data, leading principle components of variables in these vectors never explain more than 40% total variation (40% for bilateral net exports as the highest, 23% for bilateral changes in net bank claims as the lowest). This exploratory analysis provides suggestive evidence favoring the identification condition.

If the identifying condition is indeed violated, then my estimate of the γ parameter is likely to be downward biased. Adversarial shocks under tumultuous market conditions suppressing *all* bilateral trades and portfolio investments (the observables) tend to be associated with dollar shortages, boosting demands for spot dollars (via synthetic dollar funding) and dampening the need for forward dollars. If we interpret the unobservables absorbed by ℓ as forward dollar demands due to liquidity needs, the instrument constructed using $\boldsymbol{x}_{-i}$ will be positively correlated with ℓ_i in equation (21): they both drop in bad times. Estimates of $-\gamma$ (a negative object in theory) will be inflated by the instrument, which is equivalent to downward biased γ estimates.

Table 9 reports the second-step demand estimation results. In these estimations, I normalize arbitrageurs' capital to one at the beginning of the sample (January 2009). The key parameter of interest is γ. The OLS estimation of γ is negative, suggesting that this simple approach is mired by unobservable demand drivers. IV estimations yield γ estimates of around 1.4. Weak IV test statistics for the first-stage regression exceed theory cutoffs calculated following Stock and Yogo (2005).

Interpreting the number $\gamma = 1.4$ relies on estimates of $\overline{\boldsymbol{\beta}}$, the components of which are significantly positive for net purchases of long-term securities (only long-term bonds, not equities) and net exports. This finding itself is intuitive. Higher (US) net exports indicate that US exporters expect more foreign-currency receivables. To hedge these cash flows against currency risk, they sell foreign currencies forward in exchange for dollar. As a result, increased net exports indicate higher forward dollar demands. Similar reasoning apply to net foreign asset purchases which generate foreign-currency denominated cash flows (and capital gains) in the future. The coefficient is around five for net long-term bond purchases, which is 3.6 times of γ. This suggests that one basis point increase in the CIP deviations is equivalent to $1/3.6 \approx 0.28$ billion decrease in this variable in terms of impacts on hedging demands. Similarly, the coefficient for net exports is around ten ($\approx 7 \times \gamma$). Thus, one basis point increase in the CIP deviations tends to have the same impact on hedging demands as $1/7 \approx 0.14$ billion decrease in net exports.[37]

5 Quantitative Analysis

5.1 Model-implied CIP deviations

With estimates of the function $S_0(\cdot)$, α_t (determined by the vector δ), $\gamma_{0,t} = [\gamma_{01,t}, \dots, \gamma_{0n,t}]^\top$ (each element determined by $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n$ respectively), and the parameter γ, the equilibrium condition (16) becomes a pricing system: at time-t, CIP deviations $\boldsymbol{b}$ solves

$$\frac{r_t \partial S_0 \left(\alpha_t \boldsymbol{w}_t^\top \boldsymbol{b}\right)}{\partial \boldsymbol{b}} k_t = \gamma_{0,t} - \gamma \boldsymbol{b}, \tag{22}$$

where $\boldsymbol{w}_t$ contain weights calculated from FX derivatives trading volumes, k_t is measured by market equity of the 49 FX dealer banks. I solve for model-implied $\boldsymbol{b}$ each month from this equation and compare it with data.

[37] Since I do not rule out correlations between $\boldsymbol{x}_{it}$ and ℓ_{it} in equation (21), $\boldsymbol{\beta}$ estimates cannot be treated as causally identified. These results should be interpreted with caution. Calculation here may be illustrative, but can at least help better understand the model.

Denote by $\hat{b}$ the model-implied CIP deviations for a specific currency and by b the true data (one-year CIP deviations measured using currency swap rates). Figure 7 contrast CIP deviations solved from equation (22) against data. Overall, the model-implied CIP deviations track the data well. Panel (A) of Table 10 reports the means and standard deviations of b and $\hat{b}$ as well as their correlations. Sample periods are January 2009 to December 2019. Overall, moments of model-implied CIP deviations closely track ones from the data for G6 currencies.

If we decompose the variance of observed data b as

$$\sigma^2[b] = \mathrm{Cov}[b, b] = \mathrm{Cov}[b, \hat{b}] + \mathrm{Cov}[b, b - \hat{b}],$$

the ratio $\mathrm{Cov}[b, \hat{b}]/\sigma^2[b]$ measures the fraction of total variance in the data that the model accounts for. This quantity is equivalent to the slope coefficient of the following regression

$$\hat{b}_t = \beta_0 + \beta b_t + \varepsilon_t,$$

The last two columns of Table 10 report estimates of β in this regression and their standard errors. For euro, yen, pounds, and Canadian dollars, model-implied CIP deviations account for at least 57 percent of total variation in the data. For Australian dollar and Swiss Franc, the model-implied CIP deviations explain over 30 percent of observed variation. The relative poor performance for CHF is mainly due to low variation in the model-implied quantities. The correlation between b and $\hat{b}$ is 0.59 for CHF but the variance of $\hat{b}$ is 46 percent lower. Overall, the model explains around 57 percent of variation in one-year CIP deviations of G6 currencies.

Out-of-sample analysis: sample splitting. I repeat the two-step model estimation exercise using the 2009-2015 subsample, and treat the 2016-2019 subsample as testing data. Adopting a common "trick" facing the bias-variance trade-off (when performing out-of-sample prediction tasks), I choose a more parsimonious hedging demand specification, which only includes net exports and net bond purchases. Using parameters estimated from the first subsample, I solve for CIP deviations according to equation 22. Figure 8 shows the out-of-sample prediction results. Levels of predictions align well with the data in the testing sample.

Out-of-sample analysis: additional currencies. I further check model performance by applying it to four currencies *not* used in the first-step model estimation: the Swedish krona (SEK), Norwegian krone (NOK), New Zealand dollar (NZD) and Hong Kong dollar (HKD). As an approximation, when solving for equilibrium CIP deviations from equation (22) for these new currency pairs, I ignore all off-diagonal elements in the partial differentiation on the left-hand side. I use $\overline{\boldsymbol{\beta}}$ estimates from Table 9 to computer their hedging demands (instead of finding $\boldsymbol{\beta}_i$ for each currency). I compare model outcomes with data in Figure 9. Although no information regarding these currencies is used for estimation, CIP deviations solved from the model still align well with data. Panel (B) of Table 10 compares moments for the model-implied ones with data and repeat the regression analysis above.

The model tracks data moments well. On average, model-implied quantities explain over 30 percent variation in the data.

Restoring CIP deviations back to their pre-crisis levels. With equation (22), we can investigate counterfactual CIP deviations when arbitrageurs are facing tighter or loser financial constraints. I conduct this exercise via replacing α_t in equation (22) by $(c_\alpha \alpha_t)$ and resolve for equilibrium CIP deviations. A larger constant c_α indicates loser financial constraints. Table 11 reports time-series average of counterfactual CIP deviations as well as their standard deviations for different c_α. One particularly interesting observation from Table 11 is that loosening the financial constraints by allowing for 2.5 times larger arbitrage legs (recalling the illustration in Figure 3) can restore the post-crisis CIP deviations back to their pre-crisis levels (of around five basis points).

5.2 Shapley-value decomposition of the model-implied CIP deviations

To determine the relative contribution of (the dynamics of) financial constraints (α_t), hedging demands $\gamma_{0,t}$, and arbitrageurs' capital (k_t) to time-series variation in CIP deviations, I adapt a Shapley decomposition (see Shorrocks (2013) for its application in linear models) to the equilibrium pricing function. For each of the three forces, Shapley decomposition determines its marginal contribution to total variation in model-implied CIP deviations. This decomposition scheme is especially useful as the three economic forces interact with each other to determine equilibrium CIP deviations nonlinearly through equation (22). Conceptually, the three economic forces are teammates who cooperate on a task – producing variation in $\boldsymbol{b}$. The Shapley decomposition calculates their "wages" for finishing the task in an efficient, fair, and easy-to-interpret manner.

I begin by adapting the Shapley decomposition to my equilibrium model. Equation (22) defines an implicit function $\boldsymbol{b} = L(\alpha, k, \gamma_0)$ that maps the three variables to the equilibrium CIP deviations. For variable $v \in \{\alpha, k, \gamma_0\}$, I compute

$$I_v = \sum_{V \subset \{\alpha, k, \gamma_0\} \setminus \{v\}} \frac{|V|}{6} \left\{ \sigma^2[L(V, v)] - \sigma^2[L(V, \overline{v})] \right\},$$

where $\sigma^2[L(V, v)]$ denotes the variance of counterfactual CIP deviations calculated from the implicit function, holding $\{\alpha_t, k_t, \gamma_{0,t}\} \setminus \{V, v\}$ constant (as its sample average) while allowing both v and variables in V to vary; $\sigma^2[L(V, \overline{v})]$ denotes the variance calculated similarly holding both $\{\alpha_t, k_t, \gamma_{0,t}\} \setminus \{V, v\}$ and the variable of interest v constant (only variables in V are allow to change across time). For each v, the identity sums across all configurations excluding itself. Under this decomposition scheme, the variance of model-implied CIP deviations satisfies

$$\sigma^2[\hat{b}] = I_\alpha + I_k + I_{\gamma_0}.$$

Of note, for the vector γ_0, when computing counterfactual CIP deviations of currency i, only its ith element is held constant when needed.

Table 12 reports the fraction of variation in model-implied CIP deviations ($I_v/\sigma^2[\hat{b}]$) that can be attributed to each of the three drivers. On average, financial constraints are responsible for 46.4 percent of variation in model-implied CIP deviations. Hedging demands and arbitragers' capital explain the other 38.0 and 15.6 percents.

For variation in the data, consider the following equation

$$\sigma^2[b] = \sigma^2[\hat{b}] + \mathrm{Cov}[b - \hat{b}, \hat{b}] + \mathrm{Cov}[b, b - \hat{b}].$$

Since $\sigma[\hat{b}]/\sigma[b] \approx 1$ for most currencies according to Table 10, ratios above also approximate the fraction of variation in the data that can be attributed to each of the three economic forces.

These impacts differentiate across currencies. For euro and yen, the dynamics of financial constraints plays a crucial role in driving CIP deviations, accounting for 60-70 percent CIP deviations in the model. For commodity currencies including Canadian dollars and Australian dollars, hedging demands account for approximately 70 and 40 percent variation respectively. Arbitrageurs' capital dynamics exerts substantial impacts (30 percent) only on pound-dollar CIP deviations.

To further investigate the dynamics of variance attribution, I perform the Shapley decomposition on a four-year rolling-window basis. Figure 10 presents the results. The most striking pattern from plots in Figure 10 is that arbitrageurs' capital can stabilize the CIP basis when financial constraints or hedging demands exert disproportionately large impacts. That is, under the counterfactual settings of holding arbitrageurs' capital constant, fluctuations in CIP deviations can increase. For example, in 2013-2014, Canadian dollar basis is overwhelmingly driven by hedging demands. If arbitrage capital remains constant, (counterfactual) variation in Canadian dollar CIP deviations would double.

One limitation to the Shapley decomposition due to the fact that arbitrageurs' capital is endogenously determined according to equation (15). Thus counterfactual CIP deviations lead to alternative dynamics of k_t, the variation of which further generates feedbacks to the equilibrium basis. I do not account for this interaction in my current decomposition exercise. Failing to do so may exaggerate influences of arbitrageurs' capital. A potential channel is that higher CIP deviations due to relatively low levels of (contemporaneous) k_t help replenish arbitrageurs' capital in the future, enabling arbitrageurs to better absorb future financial and hedging demand shocks.

5.3 The shape of financial constraints

The (basline) shape of financial constraints, namely $\mathcal{C}_0$, can be recovered from estimates of the function $S_0(x)$ as follows

$$\bigcap_{0<\theta<\pi/2} \left\{(x, y) : x + y \tan\theta \leq S_0(\tan\theta)\right\}.$$

Intuitively, $\mathcal{C}_0$ is a set containing all points "inside" the envelope of half planes $x + y\tan\theta \le S_0(\tan\theta)$ for varying θ. Layering the half planes will unveil the shape of of financial constraints, a procedure similar to tomography: the shape of an object can be reconstructed from its shadows when light beams shine on it from many different angles. angles.[38]

One particularly interesting exercise would be figuring out how the baseline shapes of financial constraints morph across time. This shape-shifting variation can capture additional dynamics of financial constraints beyond the series α_t. To make progress, I reestimate model (19) for subsample periods of 2009-2013 and 2015-2019 and compare the recovered shape estimates.

Figure 11 presents the estimated S_0 functions as well as the recovered sets $\mathcal{C}_0$. The top and bottom panels correspond to results for 2009-2013 and 2014-2019 respectively. White areas enclosed by blue half planes are the sets $\mathcal{C}_0$. The x-axis corresponds to π_0 (fractions of equity capital deployable to support routine business) and y-axis is for π (arbitrage positions). Shapes of these sets contain important information regarding arbitrageurs' internal capital allocation decisions. Let us shift our focus to the bottom right corner of $\mathcal{C}_0$ in Panel (A). The pattern suggests that, in 2009-2013, arbitrageurs can build arbitrage positions that are almost three times of their equity capital without the need to curtail other investment positions. If arbitrage positions are four times larger than the equity capital, their routine investments will shrink about 10 percent. Going beyond this level, increased π leads to sharp decreases in π_0 and the response is almost linear, indicating pronounced balance sheet costs.

Panel (B) of Figure 11 suggests that during 2014-2019, the balance sheet space becomes more costly. Arbitrage positions quickly translate into downsized routine investments. For an arbitrage position that is five times of arbitrageurs' equity capital, the size of normal business position (π_0) is reduced by more than one half (compared with 10 percent during 2009-2013). A hard leverage cap of around seven emerges in this period. This outcome appears to be consistent with the fact that the supplementary leverage ratio (SLR) requirement was finalized in the third quarter of 2014.

6 Conclusion

Most existing limits-to-arbitrage models lack the potential to be mapped to data directly, thus the valuable insights they offer are hard to quantify. This paper attempts to partially bridge the gap by developing a quantitative model of limited arbitrage with a special focus on deviations from covered interest rate parity (CIP) conditions.

The model and its estimation methods can be a useful framework for understanding other "anomalous" pricing phenomena in today's financial markets, such as the IOER-RRP arbitrage (interest rates

[38]Rigorously speaking, the set "$\mathcal{C}_0$" recovered from S_0 using this procedure is the two-dimensional generator $\mathcal{C}_0^{2D}$ under Assumption 7. Readers may revisit Figure 4 for illustration. I will use the two notations interchangeably here. I also restrict the range of θ such that the recovered set is in the first quadrant. According to Assumption 7, $\mathcal{C}_0$ is symmetric to the horizontal axis, thus its shape in the fourth quadrant is trivial.

on excess reserves being greater than the over overnight reverse repo rates), the CDS-bond basis (the difference between credit spreads and credit default swap rates of the same bond), and negative swap spreads (thirty-year Treasury yields exceeding the corresponding swap rates). Common research questions arise in response to these phenomena. For example, who are the main arbitrageurs in these markets? What types of constraints they face (that are binding)? What are the main drivers of demands for the involved derivatives contracts? What explains time-series variation of the underlying arbitrage opportunities? The current paper illustrates how to use the framework to answer such questions.

My main contribution is to combine potential drivers of price dislocations such as hedging demands, financial constraints, and arbitrageurs' capital in a parsimonious equilibrium model. The model is flexible enough to incorporate existing knowledge about these economic forces and estimate their influences on asset prices (and their deviations from frictionless benchmarks). The key innovation is a general specification of financial constraints, and the theoretical and econometric tools developed for unveiling their shapes and capturing their dynamics.

References

Acharya, V. V., L. A. Lochstoer, and T. Ramadorai (2013). Limits to arbitrage and hedging: Evidence from commodity markets. *Journal of Financial Economics 109*(2), 441–465.

Adrian, T. and H. S. Shin (2014). Procyclical leverage and value-at-risk. *Review of Financial Studies 27*(2), 373–403.

Akram, Q. F., D. Rime, and L. Sarno (2008). Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics 76*(2), 237–253.

Akram, Q. F., D. Rime, and L. Sarno (2009). Does the law of one price hold in international financial markets? Evidence from tick data. *Journal of Banking and Finance 33*(10), 1741–1754.

Amiti, M., P. McGuire, and D. E. Weinstein (2019). International bank flows and the global financial cycle. *IMF Economic Review 67*(1), 61–108.

Andersen, L., D. Duffie, and Y. Song (2019). Funding value adjustments. *Journal of Finance 74*(1), 145–192.

Atkeson, A. G., A. d'Avernas, A. L. Eisfeldt, and P.-O. Weill (2019). Government guarantees and the valuation of American banks. *NBER Macroeconomics Annual 33*(1), 81–145.

Augustin, P., M. Chernov, L. Schmid, and D. Song (2020). The term structure of CIP violations. Technical report, National Bureau of Economic Research.

Avdjiev, S., W. Du, C. Koch, and H. S. Shin (2019). The dollar, bank leverage, and deviations from covered interest parity. *American Economic Review: Insights 1*(2), 193–208.

Baba, N. and F. Packer (2009). Interpreting deviations from covered interest parity during the financial market turmoil of 2007–08. *Journal of Banking and Finance 33*(11), 1953–1962.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Bertsekas, D. P. (2009). *Convex Optimization Theory*. Athena Scientific Belmont.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Calomiris, C. W. and D. Nissim (2014). Crisis-related shifts in the market valuation of banking activities. *Journal of Financial Intermediation 23*(3), 400–435.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cvitanić, J. and I. Karatzas (1992). Convex duality in constrained portfolio optimization. *Annals of Applied Probability 2*(4), 767–818.

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics 80*(4), 549–560.

Du, W., A. Tepper, and A. Verdelhan (2018). Deviations from covered interest rate parity. *Journal of Finance 73*(3), 915–957.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science 11*(2), 89–121.

Frenkel, J. A. and R. M. Levich (1975). Covered interest arbitrage: Unexploited profits? *Journal of Political Economy 83*(2), 325–338.

Frenkel, J. A. and R. M. Levich (1977). Transaction costs and interest arbitrage: Tranquil versus turbulent periods. *Journal of Political Economy 85*(6), 1209–1226.

Gabaix, X. and M. Maggiori (2015). International liquidity and exchange rate dynamics. *Quarterly Journal of Economics 130*(3), 1369–1420.

Gârleanu, N. and L. H. Pedersen (2011). Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies 24*(6), 1980–2022.

Gârrleanu, N., L. H. Pedersen, and A. M. Poteshman (2008). Demand-based option pricing. *Review of Financial Studies 22*(10), 4259–4299.

Greenwood, R. and D. Vayanos (2010). Price pressure in the government bond market. *American Economic Review 100*(2), 585–90.

Gromb, D. and D. Vayanos (2018). The dynamics of financially constrained arbitrage. *Journal of Finance 73*(4), 1713–1750.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica 50*(1), 213–224.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–70.

Ivashina, V., D. S. Scharfstein, and J. C. Stein (2015). Dollar funding and the lending behavior of global banks. *Quarterly Journal of Economics 130*(3), 1241–1281.

Jermann, U. (2020). Negative swap spreads and limited arbitrage. *Review of Financial Studies 33*(1), 212–238.

Klingler, S. and S. Sundaresan (2019). An explanation of negative swap spreads: Demand for duration from underfunded pension plans. *Journal of Finance 74*(2), 675–710.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Kondor, P. (2009). Risk in dynamic arbitrage: the price effects of convergence trading. *Journal of Finance 64*(2), 631–655.

Kondor, P. and D. Vayanos (2019). Liquidity risk and the dynamics of arbitrage capital. *Journal of Finance 74*(3), 1139–1173.

Kyle, A. S. and W. Xiong (2001). Contagion as a wealth effect. *Journal of Finance 56*(4), 1401–1440.

Longstaff, F. (2004). The flight-to-liquidity premium in u.s. treasury bond prices. *Journal of Business 77*(3), 511–526.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies 25*(12), 3457–3489.

Lucas, R. E. (1967). Adjustment costs and the theory of supply. *Journal of Political Economy 75*(4), 321–334.

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

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–887.

Molchanov, I. and F. Molinari (2018). *Random Sets in Econometrics*, Volume 60. Cambridge University Press.

Puriya, A. and F. Bräuning (2021). Demand effects in the FX forward market: Micro evidence from banks' dollar hedging. *Review of Financial Studies 34*(9), 4177–4215.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality 11*(1), 99.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge University Press.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics 10*(4), 1040–1053.

Taylor, M. P. (1987). Covered interest parity: a high-frequency, high-quality data study. *Economica 45*(216), 429–438.

Vayanos, D. and J.-L. Vila (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica 89*(1), 77–112.

Xu, C. (2020). Reshaping global trade: the immediate and long-run effects of bank failures. Technical report, Available at SSRN 3710455.

Table 1: Summary statistics of one-year CIP deviation measures for G6 currencies against the dollar.

	currency swap rates					forward-OIS bases				
	mean	s.d.	median	min	max	mean	s.d.	median	min	max
EUR	−28.59	16.22	−26.60	−107.75	−17.00	−37.65	19.29	−35.97	−81.87	−23.24
JPY	−34.43	14.91	−30.75	−82.38	−23.00	−53.06	22.50	−53.28	−109.38	−33.54
GBP	−9.93	11.43	−7.62	−77.07	−1.88	−13.49	12.93	−10.05	−55.89	−3.64
CAD	−10.56	10.98	−11.50	−32.88	−3.60	−8.73	12.20	−4.57	−74.27	−0.59
AUD	14.35	6.67	13.50	−4.12	18.90	13.51	14.43	13.93	−53.19	21.04
CHF	−26.52	13.10	−24.75	−80.75	−16.00	−54.21	23.53	−50.18	−102.76	−34.72

Table 2: Predictive regressions: book equity and market equity returns of global dealer banks on one-year basis swap rates

This table presents results from the following panel regressions

$$\frac{1}{\tau}\text{return}_{i,t+\tau} = \alpha_i + \beta \bar{b}_t + \varepsilon_{i,t+\tau},$$

for quarterly observations. The dependent variables are one-quarter-ahead net returns on the book equity (BE), market equity (ME), or (artificially defined "returns" on) market-to-book ratio (MB) of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. Variables are collected for their holding companies. All returns are annualized (divided by $\tau = 0.25$) net ones in percentage points. The subscript i represents banks and t denotes quarters. The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year basis swap rates or forward-OIS bases for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Sample periods begin from January 2009 and end at December 2019. Specifications with and without ($\alpha_i = \alpha$ for all $i = 1, \ldots, 49$) bank fixed effects are both included. Sample periods begin from March 2009 and end at December 2019. Numbers in parentheses are Driscoll-Kraay standard errors robust to general forms of serial correlations and cross-sectional correlations among banks (Driscoll and Kraay, 1998).

	$(\text{BE}_{i,t+\tau}/\text{BE}_{i,t} - 1)\%$		$(\text{MB}_{i,t+\tau}/\text{MB}_{i,t} - 1)\%$		$(\text{ME}_{i,t+\tau}/\text{ME}_{i,t} - 1)\%$	
Panel A: CIP deviations measured by currency swap rates						
b (b.p.)	0.244	0.227	0.627	0.639	0.875	0.871
	(0.114)	(0.108)	(0.426)	(0.420)	(0.397)	(0.395)
const.	0.62		−14.59		−14.53	
	(2.88)		(10.22)		(9.66)	
Bank f.e.	✗	✓	✗	✓	✗	✓
N obs.	1713	1713	1713	1713	1713	1713
adj.-R^2 (%)	1.0	4.0	1.5	1.4	2.9	2.4
Panel B: CIP deviations measured by forward-OIS implied bases						
b (b.p.)	0.168	0.158	0.604	0.617	0.780	0.785
	(0.082)	(0.078)	(0.356)	(0.352)	(0.341)	(0.338)
const.	2.01		−15.02		−13.66	
	(2.76)		(9.28)		(8,72)	
Bank f.e.	✗	✓	✗	✓	✗	✓
N obs.	1713	1713	1713	1713	1713	1713
adj.-R^2 (%)	0.7	3.7	1.9	1.8	3.1	2.7

Table 3: Predictive regressions: quarterly returns of FX committee surveyed (FXS) dealer banks on one-year basis swap rates and placebo tests

This table presents results from the following time-series regressions:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta\bar{b}_t + \epsilon_{t+\tau},$$

for quarterly observations. The dependent variables are one-quarter-ahead value- or equal-weighted equity returns of 49 dealer banks participating FX surveys (FXS) conducted by local monetary authority at New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. Variables are collected for their holding companies. Additional placebo tests use returns from five ETFs tracking the S&P500 index (SPY), the global financial sector (IXG), the US financial sector (IYF), US broker-dealers and securities exchanges (IAI), and US insurance companies (KIE). All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as shown in the regression specification). The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year one-year basis swap rates or forward-OIS bases for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Sample periods begin from January 2009 and end at December 2019. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

ret. (p.p.)	FXS (vw)	FXS (ew)	ETF-SPY (S&P500)	ETF-IXG (Gl. Fin.)	ETF-IYF (US Fin.)	ETF-IAI (US B&D)	ETF-KIE (US Insur.)
Panel A: CIP deviations measured by currency swap rates							
$\bar{b}$ (b.p.)	1.98	1.67	0.61	1.52	1.10	1.43	1.20
	(0.87)	(0.77)	(0.44)	(0.72)	(0.67)	(0.82)	(0.76)
const.	−32.0	−26.6	3.7	−20.4	−7.4	−17.2	−7.1
	(19.7)	(17.5)	(10.4)	(16.7)	(15.7)	(20.6)	(17.7)
N obs.	44	44	44	44	44	44	44
R^2-adj. (%)	6.2	4.5	0.6	4.9	2.5	4.0	2.1
Panel B: CIP deviations measured by forward-OIS implied bases							
$\bar{b}$ (b.p.)	1.50	1.28	0.40	1.16	0.79	0.90	0.78
	(0.64)	(0.66)	(0.26)	(0.54)	(0.45)	(0.53)	(0.49)
const.	−28.4	−23.9	6.9	−17.5	−3.5	−10.4	−0.6
	(17.4)	(17.3)	(8.1)	(15.3)	(13.6)	(18.5)	(15.5)
N obs.	43	43	43	43	43	43	43
R^2-adj. (%)	13.5	10.6	1.6	11.1	5.6	4.7	3.6

Table 4: Predictive regressions: quarterly returns of FX committee surveyed (FXS) dealer banks on one-year basis swap rates and placebo tests

This table presents results from the following time-series regressions:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta \bar{b}_t + \epsilon_{t+\tau},$$

for daily and monthly observations. The dependent variables are one-quarter-ahead value- or equal-weighted equity returns of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. Variables are collected for their holding companies. Additional placebo tests use returns from five ETFs tracking the S&P500 index (SPY), the global financial sector (IXG), the US financial sector (IYF), US broker-dealers and securities exchanges (IAI), and US insurance companies (KIE). For monthly observations, five hedge fund index returns are also included: one global composite index from BarclaysHedge (BCH), four indices from Hedge Fund Research (HFR) tracking global composite, relative value arbitrage, global-macro, and macro-currency strategies. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as shown in the regression specification). The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Sample periods begin from January 2009 and end at December 2019. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

Panel A: daily observations							
ret. (p.p.)	FXS (vw)	FXS (ew)	ETF-SPY (S&P500)	ETF-IXG (Gl. Fin.)	ETF-IYF (US Fin.)	ETF-IAI (US B&D)	ETF-KIE (US Insur.)
$\bar{b}$ (b.p.)	2.46	2.25	0.53	1.93	1.27	1.47	1.30
	(0.71)	(0.71)	(0.30)	(0.62)	(0.51)	(0.65)	(0.54)
const.	−39.8	−36.4	4.9	−28.5	−10.5	−15.2	−8.7
	(13.8)	(13.4)	(7.1)	(12.4)	(10.9)	(14.5)	(12.0)
N obs.	2859	2859	2761	2761	2761	2761	2761
R^2-adj. (%)	12.0	10.3	2.3	9.6	5.9	5.6	5.3
Panel B: monthly observations							
ret. (p.p.)	FXS (vw)	FXS (ew)	ETF-SPY (S&P500)	ETF-IXG (Gl. Fin.)	ETF-IYF (US Fin.)	ETF-IAI (US B&D)	ETF-KIE (US Insur.)
$\bar{b}$ (b.p.)	2.18	1.97	0.46	1.69	1.03	1.27	1.06
	(0.79)	(0.79)	(0.36)	(0.74)	(0.64)	(0.80)	(0.72)
const.	−34.3	−30.9	6.7	−23.5	−4.8	−10.7	−3.1
	(16.3)	(15.6)	(9.2)	(16.2)	(14.6)	(19.6)	(16.1)
N obs.	132	132	132	132	132	132	132
R^2-adj. (%)	9.7	7.9	1.0	7.1	3.3	3.6	2.8
ret. (p.p.)			BCH (Gl. Com.)	HFR (Gl. Com.)	HFR (Re. Val.)	HFR (Macro)	HFR (Macro. Cur)
$\|b\|$ (b.p.)			0.27	0.21	0.17	−0.11	0.12
			(0.16)	(0.14)	(0.12)	(0.10)	(0.12)
const.			0.1	0.4	2.6	4.0	−1.5
			(4.0)	(3.5)	(2.9)	(2.6)	(2.5)
N obs.			132	132	132	132	132
R^2-adj. (%)			1.9	1.2	1.1	0.4	1.5

Table 5: Predictive regressions: quarterly returns of FX committee surveyed dealer banks on one-year basis swap rates adjusted by controls

This table presents results from the following time-series regressions:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta\bar{b}_t + \phi \cdot \text{control}_t + \epsilon_{t+\tau}$$

for daily and monthly observations. The dependent variable is the one-quarter-ahead value-weighted equity return of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as specified in the regression equation). The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Control variables include the average smoothed earnings yield (E/P) and dividend yield (D/P) for the 49 dealer banks, the effective Fed fund rate (FFR), and the CBOE volatility index (VIX). Sample periods begin from January 2009 and end at December 2019. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

ret. (p.p.)	Daily observations			Monthly observations		
$\bar{b}$ (b.p.)	2.46	1.18	1.90	2.18	1.00	1.69
	(0.71)	(0.49)	(0.57)	(0.79)	(0.55)	(0.68)
E/P		17.1			16.6	
		(3.5)			(4.4)	
D/P			3.99			3.96
			(2.33)			(2.74)
FFR		5.26	−1.34		4.53	−1.28
		(4.61)	(5.19)		(5.75)	(6.48)
VIX		0.00	2.85		0.47	3.28
		(0.70)	(1.09)		(0.96)	(1.40)
const.	−39.8	−157.7	−92.2	−34.3	−157.7	−96.9
	(13.8)	(23.8)	(23.5)	(16.3)	(29.8)	(32.1)
N obs.	2859	2859	2859	132	132	132
R^2-adj. (%)	12.0	43.1	27.2	9.7	42.4	27.8

Table 6: Testing predictions from Proposition 1: linear regressions

This table presents results from the following time-series regressions:

$$\frac{1}{\tau}\left(\frac{\text{return}_{t+\tau}}{r_t}\right) = \beta_0 + \beta_1 X_t + \psi \times \left(\frac{1}{r_t}\right) + \phi \cdot \text{control}_t + \varepsilon_{t+\tau}, \quad X_t = \frac{\overline{b}_t}{r_t} \text{ or } \left(\frac{\overline{b}_t}{r_t}\right)^2$$

using both daily and monthly observations. The notation "$\text{return}_{t+\tau}$" denotes one-quarter-ahead value-weighted equity returns of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as shown in the regression specification). The cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar is denoted by $|b|$. The effective Fed fund rate is denoted by r. The independent variable X is the time-series of either $|b|/r$ or its square $(|b|/r)^2$. Another independent variable of interest is the inverse of the effective Fed fund rate $(1/r_t)$, inspired by the capital accumulation formula in Proposition 1. Control variables include the smoothed earnings yield (E/P) averaged across the 49 dealer banks, and the CBOE volatility index (VIX). Sample periods begin from January 2009 and end at December 2019. Robust regressions use the Huber loss function to accommodate potential outliers. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

	Daily observations				Monthly observations			
ret./r	OLS		Robust		OLS		Robust	
$\overline{b}/r$	138.5		104.1		117.1		84.9	
	(109.7)		(102.3)		(113.9)		(103.0)	
$(\overline{b}/r)^2$		33.4		30.0		24.7		27.2
		(14.0)		(12.5)		(8.9)		(3.63)
$1/r$	−0.27	−0.21	−0.15	−0.11	−0.19	−0.18	−0.12	−0.12
	(0.18)	(0.11)	(0.12)	(0.06)	(0.18)	(0.10)	(0.12)	(0.07)
E/P	79.9	88.2	68.8	76.2	59.4	64.0	69.5	79.6
	(26.7)	(27.3)	(27.0)	(27.8)	(31.3)	(33.2)	(17.6)	(18.2)
VIX	0.91	−0.19	1.45	0.53	7.79	7.51	5.05	2.52
	(7.07)	(6.56)	(4.62)	(4.43)	(8.90)	(8.13)	(6.03)	(4.96)
const.	−625.9	−621.9	−545.8	−559.3	−599.3	−566.4	−599.7	−604.6
	(133.2)	(132.9)	(177.5)	(186.4)	(149.4)	(153.8)	(89.3)	(84.8)
N obs.	2859	2859	2859	2859	132	132	132	132
R^2-adj. (%)	28.7	32.6	–	–	22.5	30.3	–	–

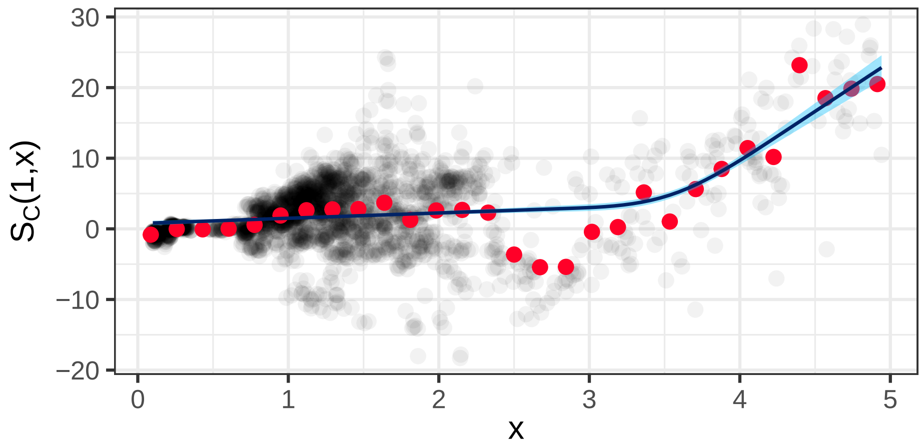
Table 7: Testing predictions from Proposition 1: semi-parametric regressions

This table presents results from the following semi-parametric regressions:

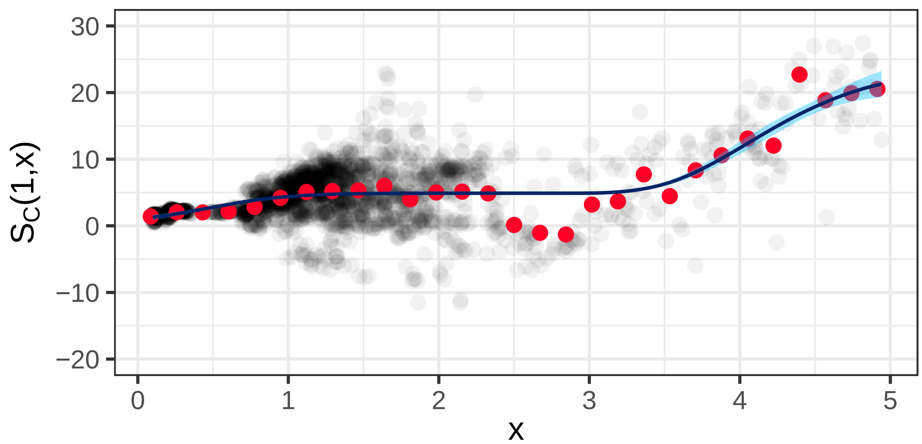
$$\frac{1}{\tau}\left(\frac{\text{return}_{t+\tau}}{r_t}\right) = S_0\left(\frac{\bar{b}_t}{r_t}\right) + \psi \times \left(\frac{1}{r_t}\right) + \phi \cdot \text{control}_t + \epsilon_{t+\tau}$$

using both daily and monthly observations. The notation "return$_{t+\tau}$" denotes one-quarter-ahead value-weighted equity returns of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as shown in the regression specification). The cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar is denoted by $|b|$. The effective Fed fund rate is denoted by r. Out of robustness concerns, only observations with $|b|/r$ falling within the their sample 5% – 95% IQR are considered. Another independent variable of interest is the inverse of the effective Fed fund rate $(1/r_t)$, inspired by the capital accumulation formula in Proposition 1. Control variables include the smoothed earnings yield (E/P) averaged across the 49 dealer banks, and the CBOE volatility index (VIX). Sample periods begin from January 2009 and end at December 2019. Semi-parametric estimation of the model uses shape-constrained B-splines basis for the functional term s. Numbers in parentheses are standard errors calculated from parametric block bootstrap procedures (that is, residuals of the fitted models are re-sampled). Block sizes are ninety for daily observations and three for monthly observations. The table also presents specification tests of whether the functional term should be included ($S_0 \equiv 0$ or not) by showing the test statistics, their (approximate) theoretical distributions, and test p-values.

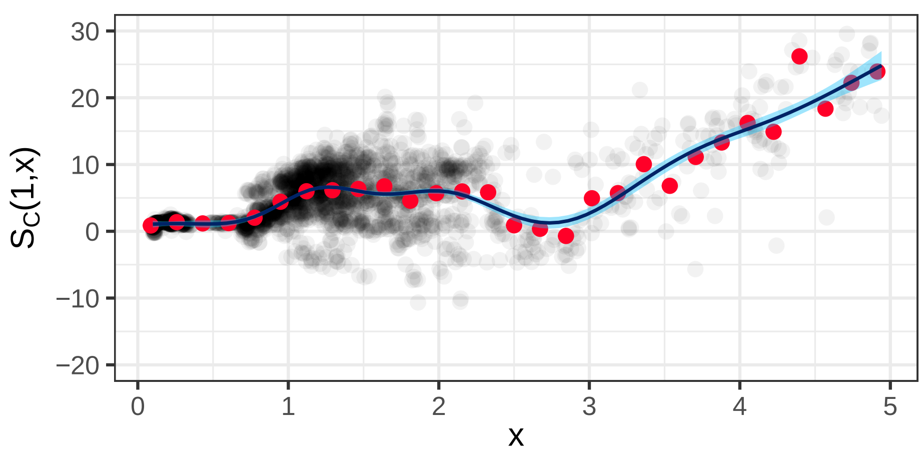
ret./r	Daily observations			Monthly observations		
$1/r$	−0.30	−0.39	−0.54	−0.22	−0.22	−0.31
	(0.10)	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)
E/P	15.8	18.0	55.6	2.3	0.66	20.7
	(38.8)	(39.7)	(37.5)	(46.9)	(48.1)	(50.9)
VIX	6.64	4.59	−2.30	9.73	10.47	6.24
	(5.95)	(6.06)	(5.27)	(7.17)	(6.88)	(7.19)
Test $H_0 : S_0 \equiv 0$ v.s. $H_1 : S_0 \not\equiv 0$						
F-stat	189	189	133	13.9	7.37	6.07
Appr. dist.	$F(3, 2538)$	$F(3, 2538)$	$F(8, 2538)$	$F(1, 116)$	$F(2, 116)$	$F(4, 116)$
p-value	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	3×10^{-4}	9×10^{-4}	3×10^{-4}
Shape constraints for $f(\cdot)$:						
increasing	✓	✓	✗	✓	✓	✗
convex	✓	✗	✗	✓	✗	✗
N obs.	2541	2541	2541	119	119	119
R^2-adj. (%)	28.0	28.7	40.2	18.7	18.6	21.6



(A) $S_{\mathcal{C}}(1, x)$: estimates under both monotonic increasing and convex constraints



(B) $S_{\mathcal{C}}(1, x)$: estimates under the monotonic increasing constraint



(C) $S_{\mathcal{C}}(1, x)$: estimates without shape constraints

Figure 5: Estimates of $S_{\mathcal{C}}(1, x)$ under different configurations

Table 8: First-step model estimation: the financial constraints

This table presents results from the following semi-parametric regressions:

$$\frac{1}{\tau}\left(\frac{\text{return}_{t+\tau}}{r_t}\right) = \left[S_0\left(\exp(\delta^\top \boldsymbol{u}_t)\frac{\overline{b}_t}{r_t}\right) + \boldsymbol{\phi}^\top \boldsymbol{v}_t\right] + \varepsilon_{t+\tau}$$

using daily observations. Model parameters are δ and $\boldsymbol{\phi}$. The functional form of $S_0(\cdot)$ is treated as unknown and also estimated. The notation "return$_{t+\tau}$" denotes one-quarter-ahead value-weighted equity returns of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as shown in the regression specification). The cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar is denoted by $\overline{b}_t$. Both volume and equal weighted results are reported. The first set of independent variables in vector $\boldsymbol{u}_t$ are the dollar index, quarterly lagged volatilities of average CIP deviations, changes in dealer banks' CDS, the TED spread (three-month dollar LIBOR rates minus the three-month Treasury bill rates), the implied volatility of euro, the VIX index, and the three-month dollar convenience yield (the RefCorp bond yield minus the treasury yield). The second set of variables in vector $\boldsymbol{v}_t$ are reciprocals of the Fed fund rates ($1/r$), the earnings yields for the 49 dealer banks (E/P), and the VIX index. Numbers in parentheses are standard errors from parametric bootstrap procedures.

	weighted by volume	equally weighted
δ: dynamics of financial constraint		
dollar index	−0.018	−0.013
	(0.009)	(0.005)
lagged vol $\overline{b}_t$ (b.p.)	−0.037	−0.036
	(0.035)	(0.029)
Δ bank CDS (%)	0.096	0.047
	(0.358)	(0.294)
TED spread (%)	3.84	2.49
	(2.12)	(1.38)
ivol euro	−0.156	−0.124
	(0.079)	(0.040)
VIX	0.059	0.040
	(0.039)	(0.030)
\$ conv. yield (%)	−0.933	−0.522
	(0.218)	(0.150)
$\boldsymbol{\phi}$: return controls		
$1/r$	−0.275	−0.376
	(0.099)	(0.111)
E/P	74.5	76.1
	(21.0)	(20.8)
VIX	0.389	1.88
	(5.523)	(5.37)
N obs.	2541	2541
Deviance R^2 (%)	44.6	43.8

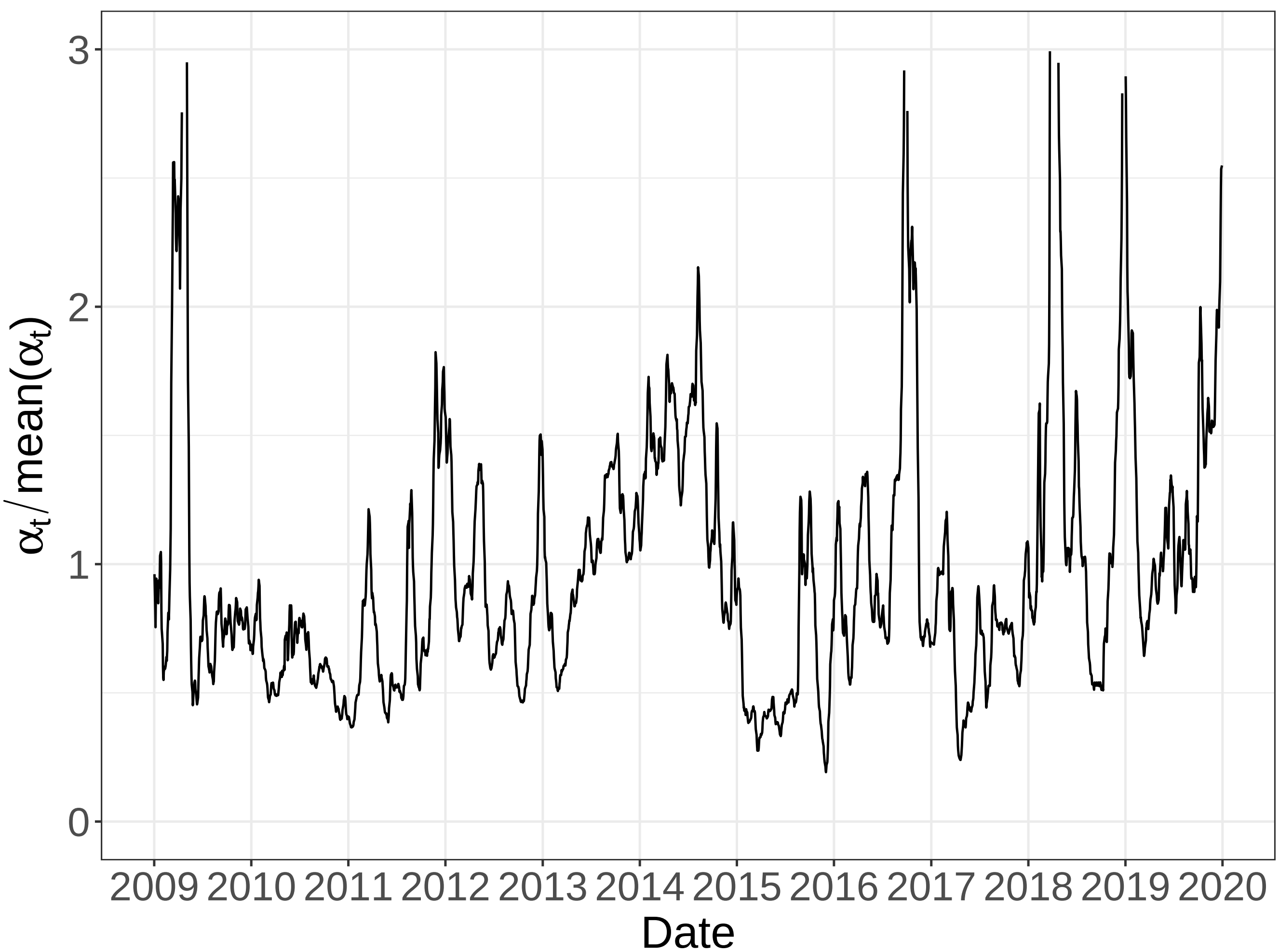


Figure 6: The dynamics of financial constraints: $\alpha_t = \exp(\delta^\top \boldsymbol{u}_t)$

Table 9: Second-step model estimation: hedging demands

This table presents results from the following panel regressions:

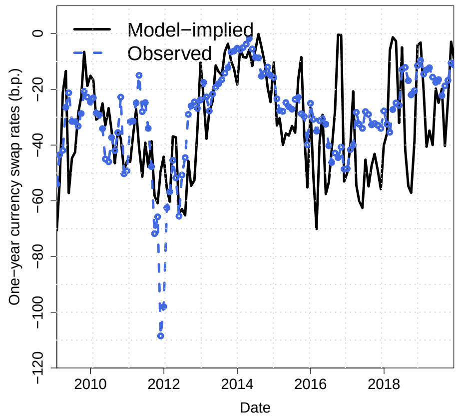
$$\pi_{it} k_t = \overline{\boldsymbol{\beta}}^{\top} \boldsymbol{x}_{it} + \sum_{j=1}^{n-1} \boldsymbol{\eta}_j^{\top} \left(I[j=i] \times \boldsymbol{x}_{jt} \right) - \gamma b_{it} + \ell_{it}$$

using monthly observations. π_{it} in the dependent variable is arbitrage positions calculated based on the first-step estimation. k_t is the total market equity of the 49 FX dealer banks, normalized to one on January 2009. The independent variables in xit include a constant one (for the intercept), bilateral net foreign direct investment, net purchases of long-term securities (sovereign and local government bonds, corporate bonds, equities), changes in net bank claims of deposits and short-term securities, and net exports (all in billions). Interest rate differentials (calculated using three-month inter-bank rates, in basis points) are also included. All "net" terms are calculated as "US minus foreign" (taking the US perspective). b_{it} stands for one-year CIP deviations for currency i at time t. The instrumental variable for b_{it} is $z_{it} = \sum_{i' \neq i} w_{i't} \boldsymbol{x}_{i't}^{(-\iota)}$, weighted average of $\boldsymbol{x}_{i't}$ ($i' \neq i$) vectors excluding the constant one (thus the "$-\iota$" superscript). The first-stage regression is then

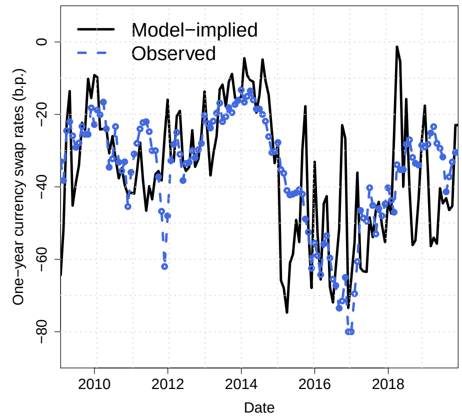
$$b_{it} = \boldsymbol{\psi}^{\top} z_{it} + \overline{\boldsymbol{\varphi}}^{\top} \boldsymbol{x}_{it} + \sum_{j=1}^{n-1} \boldsymbol{\xi}_j^{\top} \left(I[j=i] \times \boldsymbol{x}_{jt} \right) + e_{it},$$

Currencies under consideration are EUR, JPY, GBP, AUD, CAD, and CHF. The sample period is January 2009-December 2019. Numbers in parentheses are Driscoll-Kraay standard errors robust to general forms of serial correlations and correlations among currency pairs (Driscoll and Kraay, 1998).

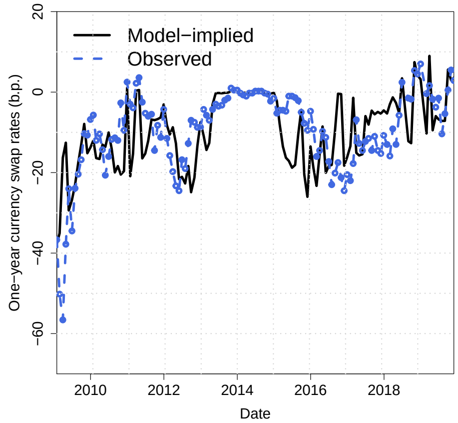
	OLS	IV	
γ	0.18	1.31	1.43
	(0.22)	(0.51)	(0.45)
Weak IV test:			
Cragg-Donald Statistic		28.7	23.5
theory cutoff (5% relative bias)		18.4	18.4
$\overline{\beta}$:			
net direct investment flows	0.92	0.47	0.37
	(0.64)	(0.79)	(0.85)
net purchase of long-term securities	−0.77	3.39	
	(0.70)	(1.59)	
• bond			5.12
			(1.64)
• equity			−1.03
			(2.15)
net change in bank claims	−0.49	−0.11	−0.10
	(0.33)	(0.33)	(0.32)
bilateral net exports	4.01	11.00	10.00
	(4.78)	(4.63)	(4.06)
$r^{\text{foreign}} - r^{\$}$ (%)	−0.03	−0.07	−0.06
	(0.03)	(0.03)	(0.03)
const.	−19.5	−45.0	−48.7
	(5.9)	(12.9)	(11.5)
N obs	784	784	784
R^2-adj. (%)	54.6	55.9	55.9



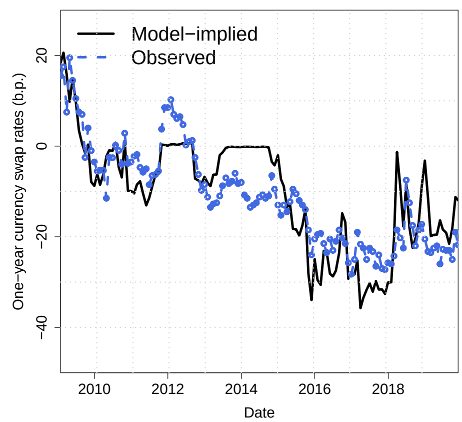
(A) EUR



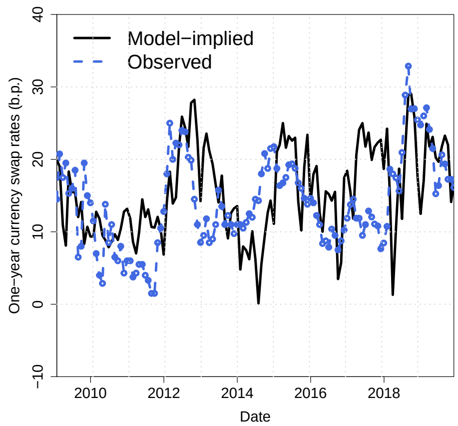
(B) JPY



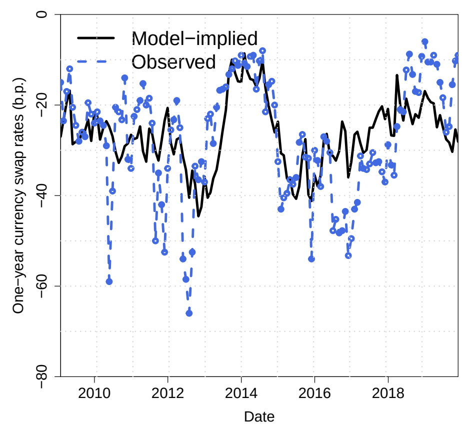
(C) GBP



(D) CAD



(E) AUD



(F) CHF

Figure 7: One-year CIP deviations for G6 currencies: model-implied and observations from currency swaps

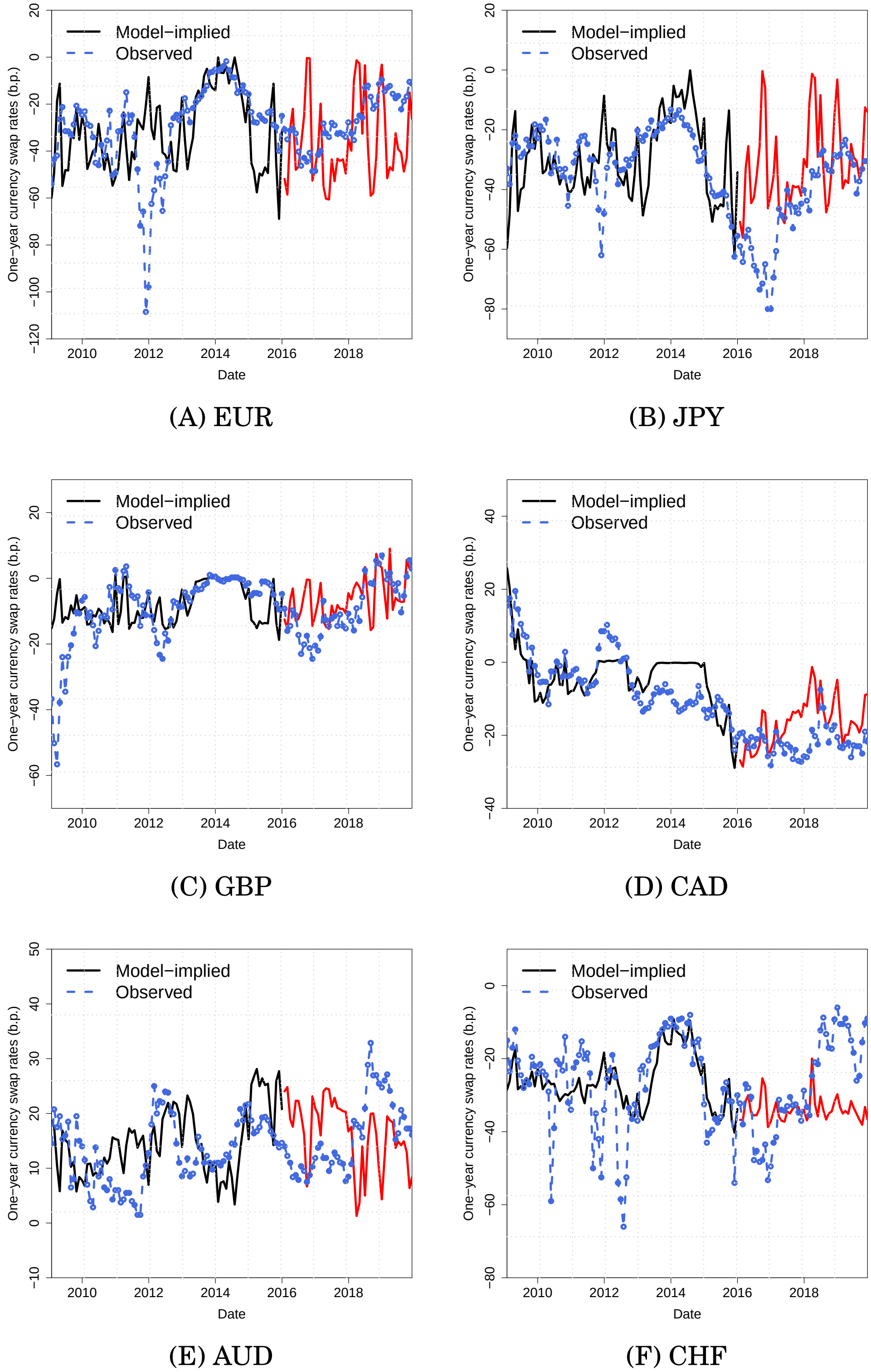
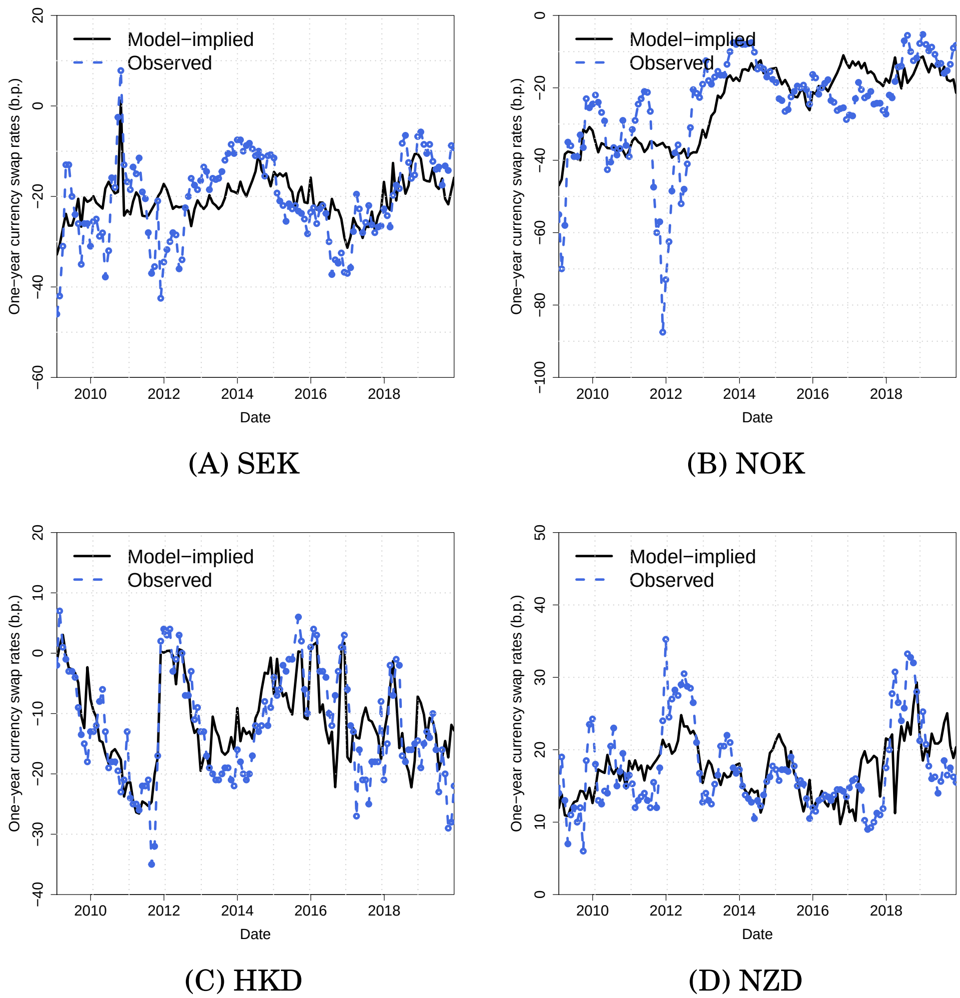


Figure 8: One-year CIP deviations for G6 currencies: model-implied (in-sample 2009-2015 in black, out-of-sample 2016-2019 in red) and observations from currency swaps



(A) SEK

(B) NOK

(C) HKD

(D) NZD

Figure 9: One-year CIP deviations for currencies not used for estimation: model-implied and observations from currency swaps

Table 10: Model-implied CIP deviations

This table documents CIP deviations solved from the estimated equilibrium equation (22), denoted by $\hat{b}$ and compares it with b, the true data (one-year currency swap rates). The parameter $\beta = \text{Cov}[\hat{b}, b]/\sigma^2[b]$ measures the fraction of variation in the data explained by the model, estimated from regressing $\hat{b}$ on b.

Currency	$\mathbb{E}[b]$	$\sigma[b]$	$\mathbb{E}[\hat{b}]$	$\sigma[\hat{b}]$	$\text{Corr}[b, \hat{b}]$	β (s.e.)	
Panel A: G6 currencies used for estimation							
EUR	−29.2	16.8	−32.4	18.6	0.56	0.62	(0.13)
JPY	−35.1	14.9	−35.4	17.8	0.61	0.73	(0.12)
GBP	−9.3	10.3	−9.7	9.0	0.59	0.52	(0.08)
CAD	−10.8	11.1	−10.6	12.2	0.83	0.91	(0.10)
AUD	14.0	6.3	15.6	6.3	0.40	0.40	(0.09)
CHF	−26.8	13.0	−26.6	7.4	0.57	0.32	(0.07)
Panel B: currencies not used for model estimation							
SEK	−20.6	9.5	−20.4	4.7	0.58	0.29	(0.06)
NOK	−24.7	14.6	−24.6	10.1	0.69	0.47	(0.08)
HKD	−12.1	9.1	−11.2	7.3	0.75	0.60	(0.07)
NZD	17.2	5.9	17.3	3.9	0.54	0.35	(0.07)

Table 11: Model-implied CIP deviations: counterfactual time-series average

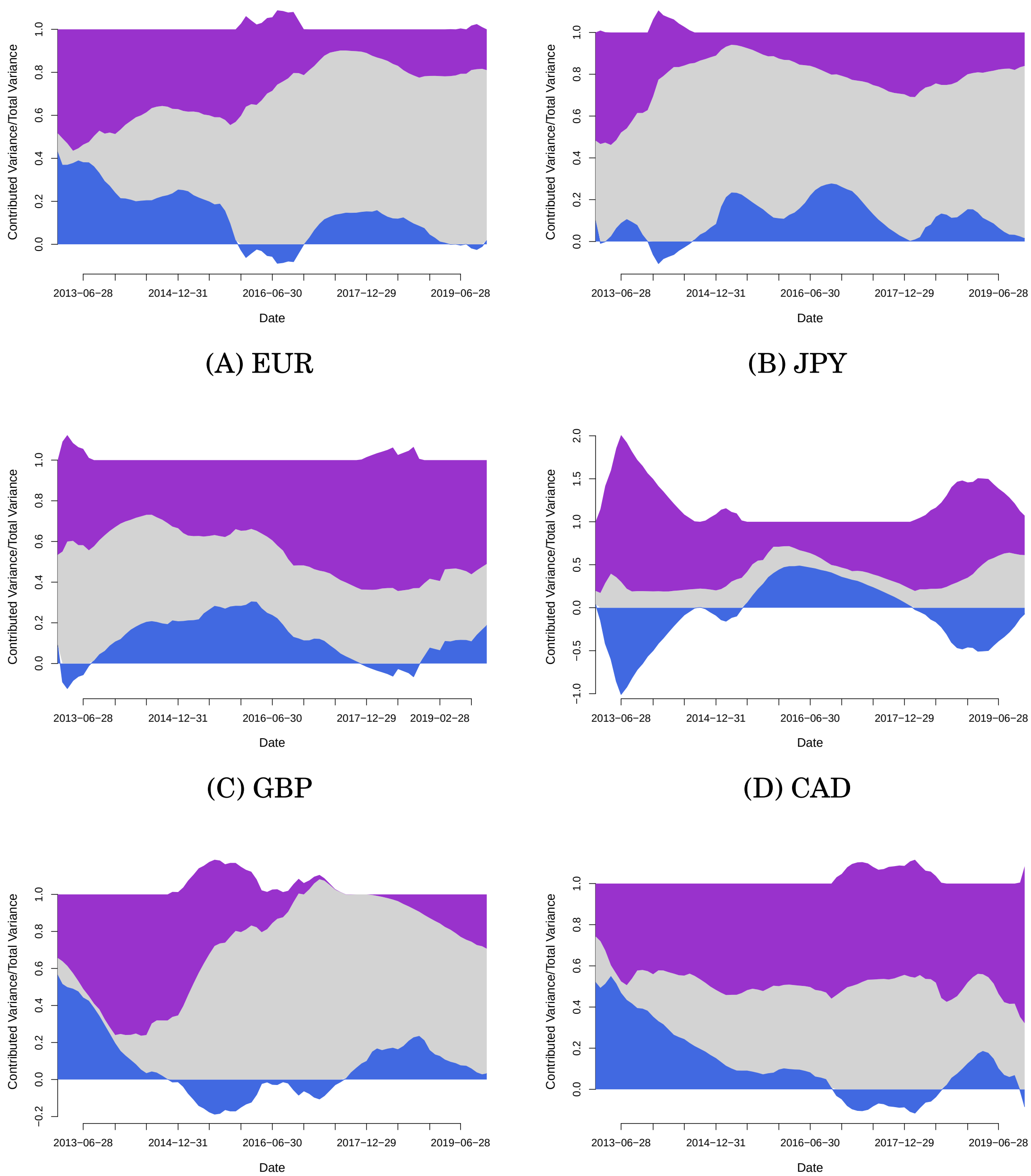
This table reports counterfactual time-series average CIP deviations from 2009 to 2019. The constant c_α relaxes ($c_\alpha < 1$) or tightens ($c_\alpha > 1$) the financial constraint. Numbers in parentheses are standard errors for these (counterfactual) sample mean statistics.

Currency	$c_\alpha = 0.5$	$c_\alpha = 1$	$c_\alpha = 1.5$	$c_\alpha = 2$	$c_\alpha = 2.5$
EUR	−51.4	−32.2	−19.2	−11.2	−6.1
	(15.3)	(17.3)	(15.1)	(11.8)	(7.1)
JPY	−50.2	−35.3	−24.1	−16.2	−8.5
	(15.3)	(17.2)	(17.1)	(14.8)	(8.2)
GBP	−14.0	−9.7	−6.5	−4.2	−2.6
	(11.2)	(9.0)	(7.5)	(6.1)	(4.5)
CAD	−13.7	−10.7	−7.5	−6.3	−4.5
	(13.7)	(12.2)	(14.3)	(9.4)	(7.3)
AUD	21.0	15.6	11.8	7.6	5.2
	(5.3)	(6.3)	(10.3)	(6.5)	(5.7)
CHF	−30.8	−26.6	−22.6	−19.0	−14.2
	(5.4)	(6.9)	(8.0)	(8.4)	(6.6)

Table 12: Shapley decomposition of model-implied CIP deviations

The table reports the full-sample Shapley decomposition results for G6 currencies. The decomposition quantifies marginal contribution from each of the three economic forces to variation in model-implied CIP deviations.

Currency	financial constraints	hedging demands	arbitrage capital
EUR	0.553	0.287	0.160
JPY	0.596	0.187	0.217
GBP	0.415	0.294	0.291
CAD	0.235	0.558	0.207
AUD	0.519	0.318	0.163
CHF	0.432	0.470	0.098
Avg.	0.458	0.352	0.190



(A) EUR

(B) JPY

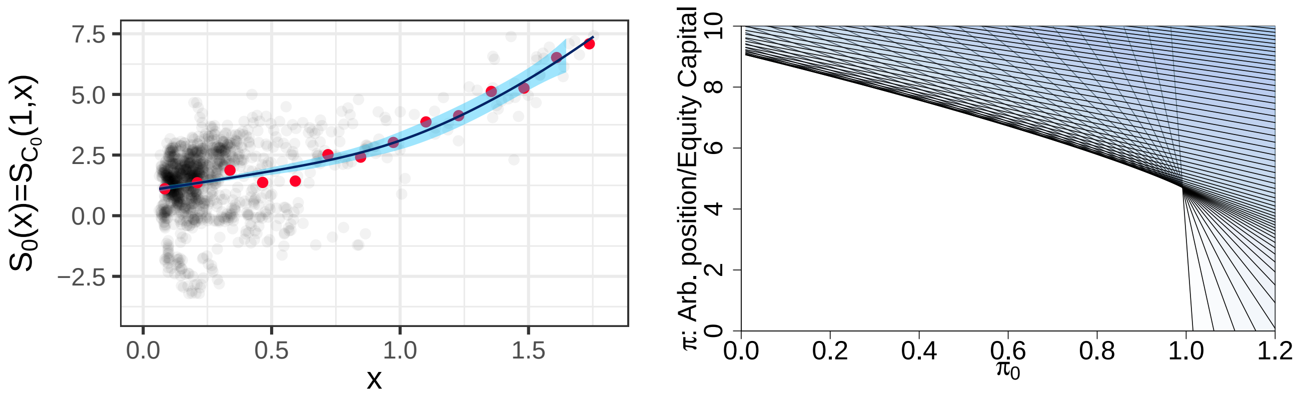
(C) GBP

(D) CAD

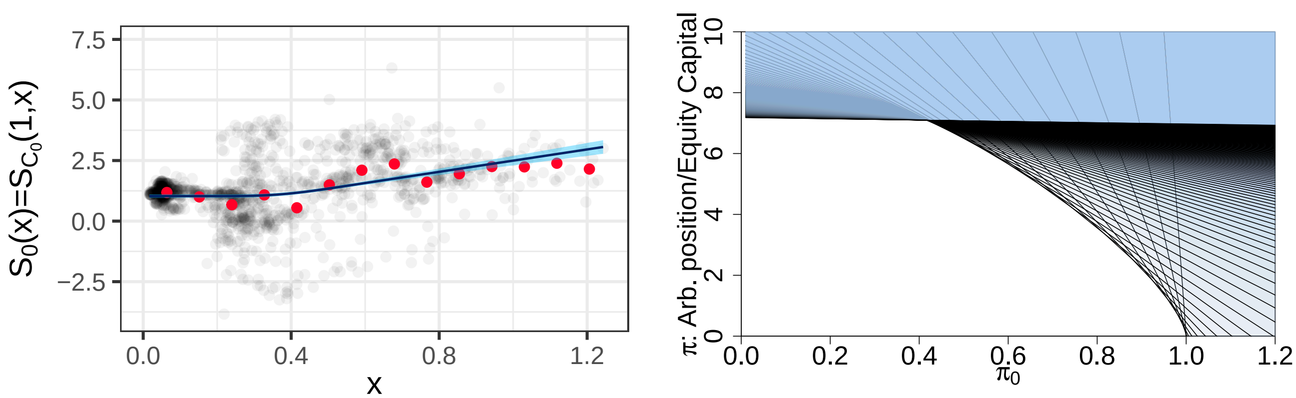
(E) AUD

(F) CHF

Figure 10: Rolling window Shapley decomposition of variation in CIP deviations (gray: financial constraints, purple: hedging demands, blue: arbitrageurs' capital)



(A): the baseline arbitrage profit function and the shape of financial constraints (2009-2013)



(B): the baseline arbitrage profit function and the shape of financial constraints (2014-2019)

Figure 11: Estimates of the baseline arbitrage profit function and shapes of financial constraints

Appendices

A Proofs

A.1 Proof of Proposition 1

I begin the proof by stating the following lemma:

Lemma A.1. *Let the pair* (π_0^*, π^*) *be such that* $\pi_0^* r + \pi^* b = S_{\mathcal{C}}(r, b)$*, then if* $b > 0$*,* $\pi^* \geq 0$*; if* $b < 0$*,* $\pi^* \leq 0$*.*

Proof of Lemma A.1. By Assumption 3, $(1, 0) \in \mathcal{C}$, as a result $S_{\mathcal{C}}(r, b) \geq r$. If $b > 0$ and $\pi^* < 0$ or $b < 0$ and $\pi^* > 0$, then $S_{\mathcal{C}}(r, b) < \pi_0 r \leq r$ (it is alway the case that $\pi_0 \leq 1$), a contradiction, thus the lemma holds.

This lemma says that when $b > 0$, arbitrageurs profit from selling dollars forward[39], i.e., $\pi^* \geq 0$ (positive forward dollar supplies). In the same vein, when $b < 0$, they will buy dollar forward, i.e., $\pi^* \leq 0$ (negative forward dollar supplies).

Now we prove Proposition 1. The arbitrageurs' optimization problem is equivalent to

$$\underset{y, s=k-y}{\text{maximize}} \quad \log(y) + \frac{1}{1+\rho}\log(k-y) \quad \text{and} \quad \underset{(\pi_0, \pi)\in\mathcal{C}}{\text{maximize}} \quad \pi_0 r + \pi b. \tag{A.1}$$

The first order condition with regard to y and s commands $s = y/(1+\rho)$. Combining this condition with $s = k - y$, we have $y = [(1+\rho)/(2+\rho)]k$ and $s = k/(2+\rho)$. By the definition of $S_{\mathcal{C}}$ and the boundedness of $\mathcal{C}$, optimal combination of π_0 and π must be such that $\pi_0 r + \pi b = S_{\mathcal{C}}(r, b) < \infty$. As a result,

$$k' = s + \pi_0 s r + \pi s b = [1 + S_{\mathcal{C}}(r, b)]\, s = \frac{1 + S_{\mathcal{C}}(r, b)}{2+\rho} k.$$

Noticing the fact that support functions are positively homogeneous of degree one (e.g., Molchanov and Molinari (2018, p. 75-76)), divide both sides of the equation above by $r > 0$ and then minus $1/r$ yield the capital accumulation equation. Substituting k by $[(2+\rho)/(1+\rho)]y$ and k' by y', we have

$$y' = \frac{1 + S_{\mathcal{C}}(r, b)}{1+\rho} y,$$

which agrees with the consumption Euler equation in the proposition.

Since $\mathcal{C}$ is convex, the support function $S_{\mathcal{C}}$ is subadditive and (thus) convex (e.g., Molchanov and Molinari (2018, p. 75-76)). It follows directly that $(k' - k)/k$ is a convex function of b.

To prove that arbitrageurs' capital gain return $(k' - k)/k$ is increasing in $|b|$, we only need to show that, with r fixed, if $b \geq 0$ ($b \leq 0$), $S_{\mathcal{C}}(r, b)$ increases (decreases) in b. Consider $b' \geq b \geq 0$,

$$S_{\mathcal{C}}(r, b') = \pi_0^{*'} r + \pi^{*'} b' \geq \pi_0^* r + \pi^* b' = \pi_0^* r + \pi^* b + \pi^*(b' - b) \geq S_{\mathcal{C}}(r, b),$$

[39] They can offer pounds for dollars to earn the favorable rate $r^{£} + b$ ($b > 0$) in cross-currency swap markets now and return dollars to reclaim pounds later, or simply buy pounds forward (with dollars) in FX swap/forward markets. Either way, they are supplying forward dollars.

where the pair $(\pi_0^{*'}, \pi_0^{*'}) \in \mathcal{C}$ maximize $\pi_0 r + \pi b'$, and the pair $(\pi_0^*, \pi_0^*) \in \mathcal{C}$ maximize $\pi_0 r + \pi b$. The last inequality above holds because $\pi^* \geq 0$ when $b \geq 0$ (Lemma A.1). Using the same notation, for $b \leq b' \leq 0$,

$$S_{\mathcal{C}}(r, b) = \pi_0^* r + \pi^* b \geq \pi_0^{*'} r + \pi^{*'} b = \pi_0^{*'} r + \pi^{*'} b' + \pi^{*'}(b - b') \geq S_{\mathcal{C}}(r, b').$$

The last inequality above follows from the fact that $\pi^{*'} \leq 0$ when $b' \leq 0$ (Lemma A.1).

Q.E.D.

A.2 Proof of Proposition 2

The arbitrageurs' optimal positions π_0^* and π^* are such that $\pi_0^* r + \pi^* b = S_{\mathcal{C}}(r, b)$. Since $S_{\mathcal{C}}$ is positively homogeneous of degree one, and is differentiable (by assumption), we can apply Euler's homogeneous function theorem (Mas-Colell, Whinston, and Green, 1995, Theorem M.B.2, p. 929), which implies that $\pi(b) = \pi^* = \partial S_{\mathcal{C}}(r, b)/\partial b$.

Then we turn to the existence and uniqueness result. Plugging the result for $\pi(b) = \partial S_{\mathcal{C}}(r, b)/\partial b$ and the expression (5) for $q(b)$ into equation (6), we have

$$b = \frac{\gamma_0}{\gamma + \dfrac{\partial S_{\mathcal{C}}(r, b)}{b \partial b} s}. \tag{A.2}$$

Noticing that, by assumption, $\pi(0) = \partial S_{\mathcal{C}}(r, 0)/\partial b = 0$ and $S_{\mathcal{C}}$ is twice differentiable,

$$\frac{\partial S_{\mathcal{C}}(r, b)}{b \partial b} = \frac{1}{b - 0}\left(\frac{\partial S_{\mathcal{C}}(r, b)}{\partial b} - \frac{\partial S_{\mathcal{C}}(r, 0)}{b \partial b}\right) = \left.\frac{\partial^2 S_{\mathcal{C}}(r, b)}{\partial b^2}\right|_{b = \hat{b} \in [0, b]}.$$

Due to the convexity of the support function $S_{\mathcal{C}}$, the condition $\partial^2 S_{\mathcal{C}}(r, b)/\partial b^2 \geq 0$ holds against any values of b for which $S_{\mathcal{C}}$ is well-defined, thus the right hand side of equation (A.2), namely $F(b)$, uniformly falls within the interval $[0, \gamma_0/\gamma]$ if $\gamma_0 \geq 0$ or $[\gamma_0/\gamma, 0]$ if $\gamma_0 < 0$. Since $S_{\mathcal{C}}$ is twice differentiable, $\pi(b) = \partial S_{\mathcal{C}}(r, b)/\partial b$ is continuous, and so is the function $F(b)$ (notice that its denominator is always positive as $\gamma > 0$). By Brouwer's fixed point theorem (Mas-Colell, Whinston, and Green, 1995, Theorem M.I.1, p. 952), the equation $F(b) = b$ admits a solution b^* in $[0, \gamma_0/\gamma]$ if $\gamma_0 \geq 0$ or $[\gamma_0/\gamma, 0]$ if $\gamma_0 < 0$, thus the existence result.

The uniqueness result follows naturally from the monotonicity of $\pi(b)s - q(b)$ in b. Since $\pi'(b)s - q'(b) = \partial^2 S_{\mathcal{C}}(r, b)/\partial b^2 s + \gamma > 0$ as long as $\gamma > 0$, $\pi(b)s - q(b)$ monotonically increases, thus the solution b^* to $\pi(b)s - q(b) = 0$ is unique.

Next, we prove the conclusion that $|b^*|$ is decreasing in the arbitrageurs' initial capital k. From the proof of Proposition 1, the arbitrageurs' saving $s = k/(2 + \rho)$, thus

$$k = \frac{(\rho + 2)(\gamma_0 - \gamma b)}{\pi(b)}. \tag{A.3}$$

The right-hand side function of b in equation (A.3), denote by $G(b)$, has a derivative

$$G'(b) = -(\rho + 2)\frac{\gamma \pi(b) + (\gamma_0 - \gamma b)\pi'(b)}{[\pi(b)]^2},$$

in which $\pi'(b) \geq 0$ for all b and $\gamma > 0$ by the assumption. When $\gamma_0 \geq 0$, $\gamma_0 - \gamma b \geq 0$ and $b \geq 0$, which implies $\pi(b) \geq 0$ from Lemma A.1. As a result, $G'(b) \leq 0$: an increase in k will leads to a smaller $b^* \geq 0$ such that $G(b^*) = k$. When $\gamma_0 < 0$, $\gamma_0 - \gamma b \leq 0$ and $b \leq 0$, indicating $\pi(b) \leq 0$ from Lemma A.1. Then $G'(b) \geq 0$: a increase in k will require a larger $b^* \leq 0$ such that $G(b^*) = k$. Summing up the conclusions, $|b^*|$ decreases in k.

Q.E.D.

A.3 Proofs (and possible generalizations) of propositions and lemmas in Section 3

Here I consider the general case: replacing the log utility with a power utility function $u(y) = (y^{1-\gamma} - 1)/(1-\gamma)$. The log utility specification is the special case when $\gamma = 1$. The arbitrageurs' problem is to maximize

$$J_t = \mathbb{E}_t \left[\int_0^\infty e^{-\rho s} u\left(y_{t+s}\right) \mathrm{d}s \right],$$

under the budget constraint (as defined in equation (11))

$$\frac{\mathrm{d}k}{k} = \pi_0 \left[r\mathrm{d}t + w(\mu - r)\mathrm{d}t + w\sigma\mathrm{d}z \right] + \boldsymbol{\pi}^\top \boldsymbol{b}\mathrm{d}t - \frac{y}{k}\mathrm{d}t,$$

and position constraint

$$(\pi_0, \boldsymbol{\pi}) \in \mathcal{C},$$

by choosing $(y, w, \pi_0, \boldsymbol{\pi})$. It is worthwhile mentioning that all time-t subscripts are omitted here. For example, the constraint $(\pi_0, \boldsymbol{\pi}) \in \mathcal{C}$ indeed represents $(\pi_{0t}, \boldsymbol{\pi}_t) \in \mathcal{C}_t$.

The Hamilton-Jacobi-Bellman (HJB) equation for arbitrageurs' optimization problem is

$$\rho J = \sup_{y, w, \pi_0, \boldsymbol{\pi}} \left\{ u(y) + \mathcal{D}J \right\}, \text{ where } (\pi_0, \boldsymbol{\pi}) \in \mathcal{C}. \tag{A.4}$$

The value function $J(k, K, \boldsymbol{s})$ is defined for the capital k of each arbitrageur, as well as the aggregate capital $K = \int_{[0,1]} k \mathrm{d}i$ of all arbitrageurs (Kyle and Xiong, 2001; Kondor and Vayanos, 2019). K is effectively an additional state variable because it determines the equilibrium arbitrage yield vector $\boldsymbol{b}$. Given that arbitrageurs are identical and of mass one, $K = k$ in equilibrium. The vector $\boldsymbol{s} \in \mathbb{R}^p$ incorporates all other state variables such as ones that determine i. the hedging demands; ii. time variation of the constraint; iii. the triplet (r_t, μ_t, σ_t) characterizing arbitrageurs' other investment opportunities. I make clear the assumptions about $\boldsymbol{s}$ as follows:

Assumption A.1. *The dynamics of $\boldsymbol{s} = (s_1, \ldots, s_p)^\top$ is written as a vector Itô process defined in a complete probability space, that is,*

$$\mathrm{d}\boldsymbol{s} = P(\boldsymbol{s})\mathrm{d}t + Q(\boldsymbol{s})\mathrm{d}\boldsymbol{z}_s,$$

where $\{\boldsymbol{z}_s\}$ is a p-dimensional vector of independent standard Brownian motions; the vector-valued function $P : \mathbb{R}^p \mapsto \mathbb{R}^p$ is such that $\sup_{\boldsymbol{s}} \|P(\boldsymbol{s})\|_2 < \infty$; the matrix-valued function $Q : \mathbb{R}^p \mapsto \mathbb{R}^{p \times p}$ is such that $Q(\boldsymbol{s})Q(\boldsymbol{s})^\top$ is positive definite with a finite dominant eigenvalue for all $\boldsymbol{s}$.

The time-varying elements of the model $\gamma_{0,t}$ (hedging demand intercepts), $\mathcal{C}_t$ (financial constraints), and (r_t, μ_t, σ_t) (investment opportunities beyond riskless arbitrage) all relate to $\boldsymbol{s}$ as follows:

i. $\gamma_{0,t} = \gamma_0(\boldsymbol{s})$ *where the mapping* $\gamma_0 : \mathbb{R}^p \mapsto \mathbb{R}^n$ *is continuously differentiable;*

ii. $\mathcal{C}_t$ *is such that its support function* $S_{\mathcal{C}_t}(x) = S_0(\boldsymbol{s}, x)$ *for all* $x \in \mathrm{dom}(S_{\mathcal{C}_t})$ *where* S_0 *is continuously differentiable in s;*

iii. $r_t = r(\boldsymbol{s})$, $\mu_t = \mu(\boldsymbol{s})$, $\sigma_t = \sigma(\boldsymbol{s})$ *where the three mappings* $r : \mathbb{R}^p \mapsto \mathbb{R}$, $\mu : \mathbb{R}^p \mapsto \mathbb{R}$ *and* $\sigma : \mathbb{R}^p \mapsto \mathbb{R}^+$ *are all continuously differentiable.*

Under the assumption above, I calculate the infinitesimal generator for the value function $J(k, K, \boldsymbol{s})$ as follows:

$$\begin{aligned}
\mathcal{D}J =& J_k \frac{\mathbb{E}\left[\mathrm{d}k\right]}{\mathrm{d}t} + \frac{1}{2} J_{kk} \frac{\mathbb{E}\left[\mathrm{d}k\mathrm{d}k\right]}{\mathrm{d}t} + J_{ks}^{\top} \frac{\mathbb{E}\left[\mathrm{d}k\mathrm{d}\boldsymbol{s}\right]}{\mathrm{d}t} + J_{kK} \frac{\mathbb{E}\left[\mathrm{d}k\mathrm{d}K\right]}{\mathrm{d}t} \\
& \underbrace{+ J_K \frac{\mathbb{E}\left[\mathrm{d}K\right]}{\mathrm{d}t} + J_s^{\top} \frac{\mathbb{E}\left[\mathrm{d}\boldsymbol{s}\right]}{\mathrm{d}t} + \frac{1}{2} \mathrm{tr}\left\{ J_{ss} \frac{\mathbb{E}\left[\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{s}^{\top}\right]}{\mathrm{d}t} \right\} + \frac{1}{2} J_{KK} \frac{\mathbb{E}\left[\mathrm{d}K\mathrm{d}K\right]}{\mathrm{d}t} + J_{Ks}^{\top} \frac{\mathbb{E}\left[\mathrm{d}K\mathrm{d}\boldsymbol{s}\right]}{\mathrm{d}t}}_{\text{constant w.r.t. } (y,\, w,\, \pi_0,\, \boldsymbol{\pi})} \\
=& J_k \left\{ k\pi_0 \left[r + w(\mu - r) \right] + k\boldsymbol{\pi}^{\top}\boldsymbol{b} - y \right\} + \frac{1}{2} J_{kk} k^2 \pi_0^2 w^2 \sigma^2 \\
& + \sum_{j=1}^{p} J_{ks_j} k\pi_0 w\sigma \frac{\mathbb{E}\left[\mathrm{d}z\mathrm{d}s_j\right]}{\mathrm{d}t} + J_{kK}^{\top} k\pi_0 w K \Pi_0 W \sigma^2 + \text{constant},
\end{aligned}$$

where $\Pi_0 = \int_{[0,1]} \pi_0 \mathrm{d}i$ and $w = \int_{[0,1]} w \mathrm{d}i$ aggregate positions π_0 and w of all arbitrageurs. In equilibrium, $\Pi_0 = \pi_0$ and $W = w$.

For the ease of exposition below, I introduce two definitions first.

Definition A.1. *(Bertsekas, 2009, Chapter. 1, p. 7, Properness of a function) A proper function* f *is one such that* $f(x) < \infty$ *for at least one* x *in its domain and* $f(x) > -\infty$ *for all* x *in its domain.*

Definition A.2. *(Bertsekas, 2009, Chapter. 1, p. 83, Conjugate functions) Consider a real-valued function* f*, the conjugate function of* f *is the function* $f^{\star}$ *defined by* $f^{\star}(y) = \sup\{x^{\top}y - f(x)\}$.

Now I begin to list and prove a set of lemmas.

Lemma A.2. *Define the indicator function for* $\mathcal{C}$*:*

$$I_{\mathcal{C}}(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C} \end{cases} . \tag{A.5}$$

$I_{\mathcal{C}}$ *is a proper closed convex function.*

Proof. First, it is always true that $I_{\mathcal{C}} \geq 0 > -\infty$, and, as long as $\mathcal{C}$ is nonempty (true by assumption), $I_{\mathcal{C}} = 0 < \infty$, for $x \in \mathcal{C}$. As a result, $I_{\mathcal{C}}$ is proper.

Second, consider the epigraph of $I_{\mathcal{C}}$, defined as $\{(x, \alpha) : I_{\mathcal{C}}(x) \leq \alpha\}$. By definition, this set is $\mathcal{C} \times [0, \infty)$, which is convex as long as $\mathcal{C}$ is convex. Thus, $I_{\mathcal{C}}$ must be convex (Bertsekas, 2009, Chapter. 1, p. 8, Definition 1.1.3).

Third, consider the set $\{x : I_{\mathcal{C}}(x) \leq \alpha\}$, which equals $\mathcal{C}$ (a closed set by assumption) when $\alpha \geq 0$ and $\emptyset$ (always closed) otherwise. Thus, $I_{\mathcal{C}}$ is a closed function. □

Lemma A.3. *The indicator function of the set $\mathcal{C}$ is such that $I_{\mathcal{C}}(x) = \sup_y \left\{x^\top y - S_{\mathcal{C}}(y)\right\}$, that is, $I_{\mathcal{C}} = S_{\mathcal{C}}^\star$.*

Proof. First, noticing that

$$S_{\mathcal{C}}(y) = \sup_{x \in \mathcal{C}} x^\top y = \sup \left\{ x^\top y - I_{\mathcal{C}}(x) \right\},$$

that is, $S_{\mathcal{C}}$ is the conjugate of $I_{\mathcal{C}}$, or simply $S_{\mathcal{C}} = I_{\mathcal{C}}^\star$.

Next, since $I_{\mathcal{C}}$ is a proper closed convex function, by the Conjugacy Theorem (Bertsekas, 2009, Chapter. 1, p. 85-86), $I_{\mathcal{C}}^{\star\star} = I_{\mathcal{C}}$, that is, the conjugate function of $I_{\mathcal{C}}^\star$ is $I_{\mathcal{C}}$ itself. Thus, $S_{\mathcal{C}}^\star = I_{\mathcal{C}}$. □

Lemma A.4. *The HJB equation of* (A.4) *under the constraint $(\pi_0, \boldsymbol{\pi}) \in \mathcal{C}$ is equivalent to*

$$\rho J = \inf_{\boldsymbol{\nu}} \sup_{y, w, \widehat{\boldsymbol{\pi}}} \left\{ u(y) + \mathcal{D}J + S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu} \right\}, \tag{A.6}$$

without any constraints, where

$$\widehat{\boldsymbol{\pi}} = \begin{pmatrix} \pi_0 \\ \boldsymbol{\pi} \end{pmatrix}$$

is a vector concatenating π_0 and $\boldsymbol{\pi}$, $\boldsymbol{\nu} = (\nu_0, \nu_1, \dots, \nu_n)^\top$ is a vector of $(n+1)$ dimensions.

Proof. Under the definition of indicator functions introduced in (A.5), the problem of (A.4) is equivalent to

$$\rho J = \sup_{y, w, \widehat{\boldsymbol{\pi}}} \left\{ u(y) + \mathcal{D}J - I_{\mathcal{C}}(\widehat{\boldsymbol{\pi}}) \right\}.$$

This equivalence is easy to understand. When $\widehat{\boldsymbol{\pi}} \in \mathcal{C}$, the optimization problem is exactly the original one. Otherwise, the indicator function penalizes the objective function so harshly that regardless how carefully the choice variables are picked, the outcome is always $-\infty$.

From Lemma A.3, $-I_{\mathcal{C}}(\widehat{\boldsymbol{\pi}}) = \inf_{\boldsymbol{\nu}} \left\{ S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu} \right\}$. Thus

$$\begin{aligned} \rho J &= \sup_{y, w, \widehat{\boldsymbol{\pi}}} \left\{ u(y) + \mathcal{D}J + \inf_{\boldsymbol{\nu}} \left\{ S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu} \right\} \right\}, \\ &= \sup_{y, w, \widehat{\boldsymbol{\pi}}} \inf_{\boldsymbol{\nu}} \left\{ u(y) + \mathcal{D}J + S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu} \right\} \\ &= \inf_{\boldsymbol{\nu}} \sup_{y, w, \widehat{\boldsymbol{\pi}}} \left\{ u(y) + \mathcal{D}J + S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu} \right\}. \end{aligned} \tag{A.7}$$

The last equation follows from the fact that the function $\{u(y) + \mathcal{D}J + S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^\top \boldsymbol{\nu}\}$ as a whole is concave in $(y, w, \widehat{\boldsymbol{\pi}})$ with fixed $\boldsymbol{\nu}$ and convex in $\boldsymbol{\nu}$ with fixed $(y, w, \widehat{\boldsymbol{\pi}})$, satisfying the saddle point property. □

Now I prove the Proposition 3. From Lemma A.4, the initial maximization problem of (A.6) leads to the following first-order condition with regard to w:

$$J_k k \pi_0 (\mu - r) + J_{kk} k^2 \pi_0^2 w \sigma^2 + \sum_{j=1}^{p} J_{ks_j} k \pi_0 \sigma \frac{\mathbb{E}\left[\mathrm{d}z \mathrm{d}s_j\right]}{\mathrm{d}t} + J_{kK} k \pi_0 K \Pi_0 W \sigma^2 = 0. \tag{A.8}$$

For elements in $\widehat{\boldsymbol{\pi}}$, the first order condition with regard to π_0 is

$$J_k k \left[r + w(\mu - r)\right] + J_{kk} k^2 \pi_0 w^2 \sigma^2 + \sum_{j=1}^{p} J_{ks_j} k w \sigma \frac{\mathbb{E}\left[\mathrm{d}z \mathrm{d}s_j\right]}{\mathrm{d}t} + J_{kK} k w K \Pi_0 W \sigma^2 - \nu_0 = 0. \tag{A.9}$$

Performing the calculation of (A.9)-[(A.8)$\times w/\pi_0$], for both sides of the two equations, we have

$$J_k k r = \nu_0. \tag{A.10}$$

For $\pi_1, \ldots, \pi_n$ in the vector $\widehat{\boldsymbol{\pi}}$, Lemma A.4 commands choosing π_i to maximize $(J_k k b_i - \nu_i)\pi_i$. As long as $J_k k b_i$ does not equal ν_i, the maximized objective function reaches infinity. Thus, in equilibrium,

$$J_k k b_i = \nu_i, \tag{A.11}$$

for all $i = 1, \ldots, n$.

Now consider the equation (A.7) shown in the proof of Lemma A.4. The initial minimization problem with regard to $\boldsymbol{\nu}$,

$$\inf_{\boldsymbol{\nu}} \left\{ u(y) + \mathcal{D}J + S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^{\top} \boldsymbol{\nu} \right\},$$

will only yield two possible outcomes: $-\infty$ or $\{u(y) + \mathcal{D}J\}$. In equilibrium, this outcome cannot be $-\infty$, thus $\boldsymbol{\nu}$ must be such that $S_{\mathcal{C}}(\boldsymbol{\nu}) - \widehat{\boldsymbol{\pi}}^{\top}\boldsymbol{\nu} = 0$, that is

$$\pi_0 \nu_0 + \sum_{i=1}^{n} \pi_i \nu_i = S_{\mathcal{C}}(\nu_0, \nu_1, \ldots, \nu_n). \tag{A.12}$$

Noticing that $S_{\mathcal{C}}$ is positively homogeneous of degree one, divide both sides of equation (A.12) by $J_k k$

$$\pi_0 \frac{\nu_0}{J_k k} + \sum_{i=1}^{n} \pi_i \frac{\nu_i}{J_k k} = S_{\mathcal{C}} \left(\frac{\nu_0}{J_k k}, \frac{\nu_1}{J_k k}, \ldots, \frac{\nu_n}{J_k k} \right),$$

and combine the result above with equation (A.10) as well as the set of equations (A.11),

$$\pi_0 r + \boldsymbol{\pi}^{\top} \boldsymbol{b} = S_{\mathcal{C}}(r, \boldsymbol{b}).$$

Proposition 3 then follows from the equation above as well as Euler's homogeneous function theorem (Mas-Colell, Whinston, and Green, 1995, Theorem M.B.2, p. 929).

Q.E.D.

For Proposition 4, I state the following generalized version (for CRRA utility functions) and then present its proof.

Proposition A.1. *Under Assumption A.1, in equilibrium, there exist a function $g(K, \boldsymbol{s}) : \mathbb{R} \times \mathbb{R}^p \mapsto \mathbb{R}$ of the aggregate capital and the state variables, such that arbitrageurs' optimal rates of consumption* *y satisfy*

$$\log\left(\frac{y}{k}\right) = \frac{1}{\gamma}\left[\log\rho - (1-\gamma)g(K, \boldsymbol{s})\right];$$

their total positions on the risky project $(\pi_0 w)$ *are*

$$\frac{1}{A(K, \boldsymbol{s})}\left(\frac{\mu - r}{\sigma^2} + \sum_{j=1}^{p}\lambda_j(K, \boldsymbol{s})\beta_j\right),$$

where

$$\beta_j = \frac{\mathrm{Cov}\left[\mathrm{d}\widetilde{r}, \mathrm{d}s_j\right]}{\mathrm{Var}[\mathrm{d}\widetilde{r}]},$$

is the regression coefficient of the changes in the j-th state variable, namely $\mathrm{d}s_j$*, regressed on the risky project return* $\mathrm{d}\widetilde{r}$*; functions* A *and* λ_j *are*

$$A(K, \boldsymbol{s}) = \gamma - (1-\gamma)\frac{\partial g(K, \boldsymbol{s})}{\partial K}K,$$

$$\lambda_j(K, \boldsymbol{s}) = (1-\gamma)\frac{\partial g(K, \boldsymbol{s})}{\partial s_j}.$$

Proof. From Lemma A.4, the maximization problem yields the following first-order-condition for y:

$$u'(y) = J_k.$$

Homogeneity of $u'(y) = y^{-\gamma}$ implies that the value function is of the following format:

$$J(k, K, \boldsymbol{s}) = \frac{1}{\rho}u\left(kG(K, \boldsymbol{s})\right), \quad \text{where } G(K, \boldsymbol{s}) = \exp(g(K, \boldsymbol{s})).$$

For the special case of log utility ($\gamma = 1$), the specification still holds and

$$J(k, K, \boldsymbol{s}) = \frac{1}{\rho}\log(k) + g(K, \boldsymbol{s}).$$

Noticing that

$$J_k = \frac{1}{\rho}u'(kG)G = \frac{1}{\rho}k^{-\gamma}G^{1-\gamma},$$

the equation $u'(y) = J_k$ is equivalent to $\rho y^{-\gamma} = k^{-\gamma}G^{1-\gamma}$. Taking logarithm and rearranging terms,

$$\log\left(\frac{y}{k}\right) = \frac{1}{\gamma}\left[\log\rho - (1-\gamma)g(K, \boldsymbol{s})\right].$$

From (A.8),

$$\pi_0 w = -\frac{J_k}{kJ_{kk}}\left(\frac{\mu - r}{\sigma^2}\right) - \sum_{j=1}^{p}\frac{J_{ks_j}}{kJ_{kk}}\underbrace{\frac{\mathbb{E}\left[(\sigma\mathrm{d}z)\mathrm{d}s_j\right]}{\mathbb{E}\left[(\sigma\mathrm{d}z)(\sigma\mathrm{d}z)\right]}}_{=\beta_j} - \frac{J_{kK}}{kJ_{kk}}K\Pi_0 W.$$

Second order derivatives of the value function are given by

$$J_{kk} = -\frac{\gamma}{\rho} k^{-\gamma-1} G^{1-\gamma} = -\frac{\gamma J_k}{k}, \quad J_{ks_j} = \frac{1-\gamma}{\rho} k^{-\gamma} G^{-\gamma} \frac{\partial G}{\partial s_j}, \quad J_{kK} = \frac{1-\gamma}{\rho} k^{-\gamma} G^{-\gamma} \frac{\partial G}{\partial K}.$$

Plugging all expressions above to the equation for $\pi_0 w$:

$$\begin{aligned}
\pi_0 w &= \frac{\mu - r}{\gamma\sigma^2} + \sum_{j=1}^{p} \frac{1-\gamma}{\gamma} \frac{\partial G}{G \partial s_j} \beta_j + \frac{1-\gamma}{\gamma} \frac{\partial G}{G \partial K} K \Pi_0 W \\
&= \frac{\mu - r}{\gamma\sigma^2} + \sum_{j=1}^{p} \frac{1-\gamma}{\gamma} \frac{\partial g}{\partial s_j} \beta_j + \frac{1-\gamma}{\gamma} \frac{\partial g}{\partial K} K \Pi_0 W \\
&= \frac{\mu - r}{\gamma\sigma^2} + \sum_{j=1}^{p} \frac{\lambda_j(K, \boldsymbol{s})}{\gamma} \beta_j + \frac{1-\gamma}{\gamma} \frac{\partial g}{\partial K} K \Pi_0 W
\end{aligned}$$

Noticing that the aggregate positions Π_0 and W equal π_0 and w respectively, then

$$\left(1 - \frac{1-\gamma}{\gamma} \frac{\partial g}{\partial K} K\right) \pi_0 w = \frac{1}{\gamma} \left(\frac{\mu - r}{\sigma^2} + \sum_{j=1}^{p} \lambda_j(K, \boldsymbol{s}) \beta_j \right),$$

that is,

$$\pi_0 w = \frac{1}{A(K, \boldsymbol{s})} \left(\frac{\mu - r}{\sigma^2} + \sum_{j=1}^{p} \lambda_j(K, \boldsymbol{s}) \beta_j \right).$$

□

Proposition 4 is the special case of the results above when $\gamma = 1$. Results collected here in Proposition A.1 have natural interpretations. The consumption-to-wealth ratio y/k is a function of the state variables $\boldsymbol{s}$ and aggregate capital K when $\gamma \neq 1$. It equals the constant ρ for the log utility case.

Arbitrageurs' demand for the risky project (proportional to their capital) is "Mertonian", both the myopic mean-variance demand and state-variable hedging demands (driven by the betas) appear when $\gamma \neq 1$. The ratio $-\lambda_j/A$ can be interpreted as the risk premium of the risky project due to its exposure to the risk factor s_j. With the log utility, $\lambda_j(K, \boldsymbol{s}) = 1$ for all $j = 1, \ldots, p$, and all hedging demands disappear.

Since the aggregate capital also becomes an endogenous state-variable, a dynamic risk-aversion function $A(K, \boldsymbol{s})$ emerges and replaces the constant relative risk-aversion parameter γ, similar to the exposition of Kondor and Vayanos (2019). With the log utility specification, the dynamic risk-aversion $A(K, \boldsymbol{s})$ equals one. Proof of Proposition 4 thus follows through.

Q.E.D.

Next I present and prove a generalized version of Proposition 5.

Proposition A.2. *In equilibrium, the arbitrageurs' capital evolves according to the following rule:*

$$\frac{\mathrm{d}k}{k} = \left[\lambda\hat{\sigma} - \frac{y}{k} + S_{\mathcal{C}}(r, \boldsymbol{b})\right] \mathrm{d}t + \hat{\sigma}\mathrm{d}z \tag{A.13}$$

where

$$\hat{\sigma} = \frac{1}{\mathrm{d}t}\mathbb{E}\left[\left(\frac{\mathrm{d}k}{k}\right)^2\right] = \frac{1}{A(K,\boldsymbol{s})}\left(\lambda + \sum_{j=1}^{p}\sigma\lambda_j(K,\boldsymbol{s})\beta_j\right);$$

$$\frac{y}{k} = \exp\left\{\frac{1}{\gamma}\left[\log\rho - (1-\gamma)g(K,\boldsymbol{s})\right]\right\};$$

functions $\lambda_j(K,\boldsymbol{s})$, $j = 1,\ldots,p$, and $A(K,\boldsymbol{s})$ are defined as in Proposition A.1; $\lambda = (\mu - r)/\sigma$ is the Sharpe ratio of the risky project available to arbitrageurs.

Proof. Plugging results from Proposition A.1 into the dynamic budget constraint of arbitrageurs, we have

$$\begin{aligned}\frac{\mathrm{d}k}{k} &= \pi_0\left[r\mathrm{d}t + w(\mu - r)\mathrm{d}t + w\sigma\mathrm{d}z\right] + \boldsymbol{\pi}^\top\boldsymbol{b}\mathrm{d}t - \frac{y}{k}\mathrm{d}t \\ &= \left(\pi_0 r + \boldsymbol{\pi}^\top\boldsymbol{b}\right)\mathrm{d}t + \pi_0 w(\mu - r)\mathrm{d}t - \frac{y}{k} + \pi_0 w\sigma\mathrm{d}z \\ &= S_{\mathcal{C}}(r,\boldsymbol{b})\mathrm{d}t + \frac{1}{A(K,\boldsymbol{s})}\left(\lambda^2 + \sum_{j=1}^{p}\lambda\sigma\lambda_j(K,\boldsymbol{s})\beta_j\right)\mathrm{d}t - \frac{y}{k}\mathrm{d}t + \frac{1}{A(K,\boldsymbol{s})}\left(\lambda + \sum_{j=1}^{p}\sigma\lambda_j(K,\boldsymbol{s})\beta_j\right)\mathrm{d}z,\end{aligned}$$

in which

$$\frac{y}{k} = \exp\left\{\frac{1}{\gamma}\left[\log\rho - (1-\gamma)g(K,\boldsymbol{s})\right]\right\}$$

in equilibrium. The proposition follows through. □

With $\gamma = 1$ (the log-utility case), functions $A = 1$ and $\lambda_j = 0$, as a result, $\hat{\sigma} = \lambda$. Also, the ratio y/k is the constant ρ in equilibrium under the log utility as in Proposition 4. Plugging these quantities back to equation (A.13), we have

$$\frac{\mathrm{d}k}{k} = \left[\lambda^2 - \rho + S_{\mathcal{C}}(r,\boldsymbol{b})\right]\mathrm{d}t + \lambda\mathrm{d}z,$$

completing the proof for Proposition 5.

Q.E.D.

I now prove Proposition 6. First, I show that it is alway the case that $k > 0$. From Proposition 5, arbitrageurs' date-t capital in equilibrium is

$$k_t = k_0\exp\left\{\int_0^t\left[\frac{1}{2}\lambda_s^2 - \rho + S_{\mathcal{C}}(r_s,\boldsymbol{b}_s)\right]\mathrm{d}s + \int_0^t\lambda_s\mathrm{d}z_s\right\},$$

which is greater than zero as long as $k_0 > 0$ (by model assumption).

Now consider the closed ball $B(0, \|\gamma_0\|_2/\gamma)$ in $\mathbb{R}^n$ and an arbitrary vector $\boldsymbol{b}$ in this ball. For any fixed r and $k > 0$, since

$$\begin{aligned}
\frac{\partial S_{\mathcal{C}}(r,\, \boldsymbol{b})}{\partial \boldsymbol{b}} &= \frac{\partial S_{\mathcal{C}}(r,\, \mathbf{0})}{\partial \boldsymbol{b}} + \frac{\partial^2 S_{\mathcal{C}}(r,\, \boldsymbol{b}^*)}{\partial \boldsymbol{b} \partial \boldsymbol{b}^\top}\boldsymbol{b} \\
&= \boldsymbol{\pi}(\mathbf{0}) + \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)\boldsymbol{b}, \\
&= \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)\boldsymbol{b} \qquad\qquad (\boldsymbol{\pi}(\mathbf{0}) = \mathbf{0} \text{ By Assumption 6})
\end{aligned}$$

for some $\boldsymbol{b}^*$ (as a function of $\boldsymbol{b}$) such that $\boldsymbol{b}^* \in B(0,\, \boldsymbol{b}) \subset B(0,\, \|\gamma_0\|_2/\gamma)$, where $\boldsymbol{H}_{\mathcal{C}} = \partial^2 S_{\mathcal{C}}/\partial \boldsymbol{b} \partial \boldsymbol{b}^\top$ defines the Hessian matrix of $S_{\mathcal{C}}$, we have $\boldsymbol{b} = (\gamma + \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)k)^{-1}\gamma_0$. Convexity of the support function $S_{\mathcal{C}}$ commands that $\boldsymbol{H}_{\mathcal{C}}$ is positive semi-definite everywhere. Let

$$\boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*) = \boldsymbol{\Gamma}^\top \mathrm{diag}(d_1, \ldots, d_n)\boldsymbol{\Gamma}, \quad d_1 \ge d_2 \ge \cdots \ge d_n \ge 0$$

be its eigen-decomposition (for $\boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)$ is real-valued and symmetric, this decomposition must exist), then

$$(\gamma + \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)k)^{-1} = \boldsymbol{\Gamma}\mathrm{diag}\left(\frac{1}{\gamma + d_1 k}, \ldots, \frac{1}{\gamma + d_n k}\right)\boldsymbol{\Gamma}^\top$$

and

$$\|\,(\gamma + \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*)k)^{-1}\gamma_0\|_2 \le \frac{1}{\gamma + d_n k}\|\gamma_0\|_2 \le \frac{1}{\gamma}\|\gamma_0\|_2.$$

Thus $(\gamma + \boldsymbol{H}_{\mathcal{C}}(r,\, \boldsymbol{b}^*(\boldsymbol{b}))k)^{-1}\gamma_0$ is a continuous mapping from $B(0,\, \|\gamma_0\|_2/\gamma)$ to itself. By Brouwer's fixed point theorem (Mas-Colell, Whinston, and Green, 1995, Theorem M.I.1, p. 952), the original equation, which finds the fixed point of this mapping, admits a solution.

Uniqueness of the solution is due to the fact that the Jocabian matrix of function $(\partial S_{\mathcal{C}}(r,\, \boldsymbol{b})/\partial \boldsymbol{b})k + \gamma \boldsymbol{b}$ is

$$J(r,\, \boldsymbol{b}) = H_{\mathcal{C}}(r,\, \boldsymbol{b}) + \gamma,$$

the determinant of which equals $\prod_{i=1}^n (d_i + \gamma) > 0$ everywhere. Thus, by the implicit function theorem (Mas-Colell, Whinston, and Green, 1995, Theorem M.E.1, p. 941-942), within the ball $B(0,\, \|\gamma_0\|_2/\gamma)$, equation $(\partial S_{\mathcal{C}}(r,\, \boldsymbol{b})/\partial \boldsymbol{b})k + \gamma \boldsymbol{b} = \gamma_0$ admits a unique solution.

Q.E.D.

I finish this section by proving Lemma 1. For $(\pi_0^*,\, \boldsymbol{\pi}^*) \in \mathcal{C}_t$ such that $S_{\mathcal{C}_t}(r,\, \boldsymbol{b}) = \pi_0^* r + \boldsymbol{\pi}^{*\top}\boldsymbol{b}$, we have that

$$S_{\mathcal{C}_t}(r,\, \boldsymbol{b}) = \pi_0^* r + \left(\frac{\boldsymbol{\pi}^*}{\alpha_t}\right)^\top (\alpha_t \boldsymbol{b}) \le S_{\mathcal{C}_0}(r,\, \alpha_t \boldsymbol{b})$$

because $(\pi_{0t}^*,\, \boldsymbol{\pi}_t^*/\alpha_t) \in \mathcal{C}_0$. For any $(\pi_0,\, \boldsymbol{\pi}) \in \mathcal{C}_0$, since $(\pi_0,\, \alpha_t \boldsymbol{\pi}) \in \mathcal{C}_t$, it must be that

$$S_{\mathcal{C}_t}(r,\, \boldsymbol{b}) \ge \pi_0 r + (\alpha_t \boldsymbol{\pi})^\top \boldsymbol{b} = \pi_0 r + \boldsymbol{\pi}^\top(\alpha_t \boldsymbol{b}).$$

Since the inequality above holds for an arbitrary pair of $(\pi_0,\, \boldsymbol{\pi}) \in \mathcal{C}_0$, $S_{\mathcal{C}_t}(r,\, \boldsymbol{b}) \ge S_{\mathcal{C}_0}(r,\, \alpha_t \boldsymbol{b})$. Combining results above, $S_{\mathcal{C}_t}(r,\, \boldsymbol{b}) = S_{\mathcal{C}_0}(r,\, \alpha_t \boldsymbol{b})$, which is the conclusion of Lemma 1.

Q.E.D.

B Algorithmic details for the first-step estimation

The statistical model is equivalent to

$$\mathbb{E}\left[y_t\right] = \boldsymbol{\lambda}^\top \boldsymbol{z}_t + S\left(x_t \alpha_t\right),$$
$$\alpha_t = \exp\left(\boldsymbol{\delta}^\top \boldsymbol{u}_t\right).$$

The semi-parametric nonlinear least square problem to solve is

$$\underset{\boldsymbol{\lambda},\boldsymbol{\delta},S(\cdot)}{\text{minimize}} \quad \sum_{t=1}^{T}\left[y_t - \boldsymbol{\lambda}^\top \boldsymbol{z}_t - S\left(x_t \alpha_t\right)\right]^2.$$

To start the algorithm, initialize $\boldsymbol{\delta}$ with a guess $\boldsymbol{\delta}^{(0)}$. At iteration i,

- Treating $\boldsymbol{\delta}^{(i)}$ as known, calculate $\alpha_t^{(i)}$. Then fit the semi-parametric model

  $$\underset{\boldsymbol{\lambda},S}{\text{minimize}} \quad \sum_{t=1}^{T}\left[y_t - \boldsymbol{\lambda}^\top \boldsymbol{z}_t - S\left(x_t \alpha_t^{(i)}\right)\right]^2$$

  to find $\boldsymbol{\lambda}^{(i)}$, $S^{(i)}(\cdot)$ and the residuals $\varepsilon_t^{(i)}$;

- Define $\hat{y}_t^{(i)} = y_t - \boldsymbol{\lambda}^{(i)\top} \boldsymbol{z}_t$, solve the following problem

  $$\underset{\boldsymbol{\delta}}{\text{minimize}} \quad L = \sum_{t=1}^{T}\left[\hat{y}_t^{(i)} - S^{(i)}\left(x_t \exp\left(\boldsymbol{\delta}^\top \boldsymbol{u}_t\right)\right)\right]^2.$$

  Specifically, consider the Taylor expansion at $x_t \alpha_t^{(i)}$

  $$\begin{aligned} L &\approx \sum_{t=1}^{T}\left[\hat{y}_t^{(i)} - S^{(i)}\left(x_t \alpha_t^{(i)}\right) - S^{(i)\prime}\left(x_t \alpha_t^{(i)}\right) x_t \left(\exp\left(\boldsymbol{\delta}^\top \boldsymbol{u}_t\right) - \alpha_t^{(i)}\right)\right]^2 \\ &= \sum_{t=1}^{T}\left[\hat{y}_t^{(i)} - S^{(i)}\left(x_t \alpha_t^{(i)}\right) + S^{(i)\prime}\left(x_t \alpha_t^{(i)}\right) x_t \alpha_t^{(i)} - S^{(i)\prime}\left(x_t \alpha_t^{(i)}\right) x_t \exp\left(\boldsymbol{\delta}^\top \boldsymbol{u}_t\right)\right]^2. \end{aligned}$$

  Define

  $$w_t = S^{(i)\prime}\left(x_t \alpha_t^{(i)}\right) x_t$$
  $$\tilde{y}_t^{(i)} = \frac{1}{w_t} \underbrace{\left[\hat{y}_t^{(i)} - S^{(i)}\left(x_t \alpha_t^{(i)}\right)\right]}_{\varepsilon_t^{(i)}} + \alpha_t^{(i)},$$

  the approximate problem becomes

  $$\underset{\boldsymbol{\delta}}{\text{minimize}} \quad L = \sum_{t=1}^{T} w_t^2 \left[\tilde{y}_t^{(i)} - \exp(\boldsymbol{\delta}^\top \boldsymbol{u}_t)\right]^2.$$

which is a weighted nonlinear least square problem. Solve the problem to get $\delta^{(i+1)}$.

- Start iteration $i+1$

The algorithm iterates the above loop until convergence.

C Microfoundation of (net) hedging demands

There are two countries, country d (domestic, the US) and f (foreign, the UK). Each country issues its own currency. We call the domestic (d) currency "dollar" and foreign (f) currency "pound". The exchange rate (pounds against dollars) at date t is E_t. In other words, one pound is exchanged for E_t amount of dollars (in practice, the GBP/USD currency pair). I assume that the dynamics of this exchange rate follows a Geometric Brownian motion:[40]

$$\frac{\mathrm{d}E}{E} = \mu_e \mathrm{d}t + \sigma_e \mathrm{d}z_e,$$

where μ_e measures the expected rate of appreciation for pounds, σ_e captures its volatility, the process $\{z^e\}$ is a standard Brownian motion.

There is a continuum of mass one identical hedgers in each country, namely d-hedgers and f-hedgers. Hedgers are exposed to currency risks due to their endowments abroad. Specifically, j-hedgers' ($j \in \{d, f\}$) endowments at time t is D_t^j from abroad, denominated in foreign currencies (i.e., D_t^f is denominated in dollars, and D_t^d is denominated in pounds). These endowments can be interpreted as cash flows from each country's Balance of Payments (BOP) items, such as export receivables, (changes in) direct or portfolio investment, as well as returns received from existing asset positions abroad. I assume that these endowments, denominated in dollars, satisfy multi-factor structures:

$$\begin{aligned} D^d E &= \boldsymbol{\lambda}_d^\top \boldsymbol{x} + \lambda_{d,0}, \\ D^f &= \boldsymbol{\lambda}_f^\top \boldsymbol{x} + \lambda_{f,0}, \end{aligned}$$

in which the vector of factors, denoted by $\boldsymbol{x}$, is a multivariate Itô process:

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{\mu}(\boldsymbol{x})\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{x})\mathrm{d}\boldsymbol{z}_x.$$

Elements in the vector $\{\boldsymbol{z}_x\}$ are standard Brownian motions. Functions $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are such that $\mathbb{E}[\mathrm{d}\boldsymbol{x}] = \boldsymbol{\mu}(\boldsymbol{x})\mathrm{d}t$ and $\mathbb{E}\left[\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}^\top\right] = \boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})^\top \mathrm{d}t$ are well defined in a complete probability space.

Hedgers of type j maximize mean-variance utilities over instantaneous wealth changes in $[t, t + \mathrm{d}t]$ denominated in their home currency

$$\mathbb{E}\left[\mathrm{d}W^j\right] - \frac{A^j}{2}\mathrm{Var}\left[\mathrm{d}W^j\right], \quad j \in \{d, f\}, \tag{C.1}$$

where $\mathrm{d}W^j$ represents these instantaneous wealth changes. The parameter A^j captures the level of risk-aversion of the type-j hedgers who face the trade-off between the mean and variance of wealth change $\mathrm{d}W^j$. Agents optimizing (C.1) can be interpreted as overlapping generations who are born at date t, manage their wealth from t to $(t + \mathrm{d}t)$, consume everything and then die at time $(t + \mathrm{d}t)$. If their preferences over final consumptions are characterized by the von Neumann-Morgenstern expected utility $\mathbb{E}[u(\cdot)]$, the risk-aversion parameter A^j in problem (C.1) can be regarded as $A^j =$

[40] Again, all time subscripts are omitted whenever there is no confusion caused.

$-u''(W^j)/u'(W^j)$.[41] I assume that

$$A^d = \frac{A^f}{E} = A,$$

just to guarantee that the risk aversion parameters are invariant against the exchange rate.

In the baseline setting, $\mathrm{d}W^d = \mathrm{d}\left(D^d E\right)$ (d-hedgers in the US repatriating pound endowments) and $\mathrm{d}W^f = \mathrm{d}\left(D^f/E\right)$ (f-hedgers in the UK repatriating dollar endowments). Hedgers cannot manage their wealth changes from time t to $(t + \mathrm{d}t)$.

Beyond the baseline setting, I allow hedgers to alter their currency risk exposures using forward contracts.[42] These contracts, signed at time t, allow hedges to exchange F units of dollars for one pound at time $(t + \mathrm{d}t)$. Taking CIP deviations as given, the forward price F satisfy

$$F \exp\left(r^f \mathrm{d}t + b\mathrm{d}t\right) = E \exp\left(r^d \mathrm{d}t\right),$$

where r^f and r^d are pound and dollar risk-free rates. For d-hedgers managing their pound exposures, they can sign a forward contract exchanging h^d pounds for $h^d F$ dollars. As a result, their wealth changes is now

$$\begin{aligned} dW^d &= \left[h^d F + \left(D^d + \mathrm{d}D^d - h^d\right)(E + \mathrm{d}E)\right] - D^d E \\ &= h^d E\left(\frac{F}{E} - 1 - \frac{\mathrm{d}E}{E}\right) + \mathrm{d}\left(D^d E\right) \\ &= h^d E\left[\left(r^d - r^f - b - \mu_e\right)\mathrm{d}t - \sigma_e \mathrm{d}z_e\right] + \mathrm{d}\left(D^d E\right). \end{aligned} \tag{C.2}$$

Choosing h^d to maximize (C.1) for $j = d$ under (C.2) yields the following first-order condition:

$$E\left(r^d - r^f - b - \mu_e\right) - A^d\left\{h^d E^2 \sigma_e^2 - \frac{E\sigma_e}{\mathrm{d}t}\mathbb{E}\left[\mathrm{d}z_e \mathrm{d}\left(D^d E\right)\right]\right\} = 0,$$

from which we solve for h^d as

$$\begin{aligned} h^d &= \frac{r^d - r^f - \mu_e}{EA^d\sigma_e^2} - \frac{b}{EA^d\sigma_e^2} + \frac{1}{E\sigma_e \mathrm{d}t}\mathbb{E}\left[\mathrm{d}z_e \mathrm{d}\left(D^d E\right)\right] \\ &= \frac{r^d - r^f - \mu_e}{EA\sigma_e^2} - \frac{b}{EA\sigma_e^2} + \frac{D^d}{\sigma_e^2 \mathrm{d}t}\mathbb{E}\left[\frac{\mathrm{d}E}{E}\left(\frac{\mathrm{d}D^d}{D^d} + \frac{\mathrm{d}E}{E} + \frac{\mathrm{d}D^d}{D^d}\frac{\mathrm{d}E}{E}\right)\right] \\ &= \frac{r^d - r^f - \mu_e}{EA\sigma_e^2} - \frac{b}{EA\sigma_e^2} + \frac{D^d}{\sigma_e^2 \mathrm{d}t}\mathbb{E}\left[\frac{\mathrm{d}E}{E}\frac{\mathrm{d}D^d}{D^d} + \sigma_e^2 \mathrm{d}t\right] \\ &= -\frac{\mu_e + r^f - r^d}{EA\sigma_e^2} - \frac{b}{EA\sigma_e^2} + \underbrace{\frac{\mathrm{Cov}\left[\mathrm{d}E/E, \mathrm{d}D^d/D^d\right]}{\mathrm{Var}\left[\mathrm{d}E/E\right]}}_{\beta_d} D^d + D^d. \end{aligned} \tag{C.3}$$

[41] This is, of course, also due to the fact that shocks to hedgers' endowment factors x and spot exchange rate return $\mathrm{d}E/E$ are all Gaussian.

[42] In practice, other forward-like derivative contracts such as foreign exchange swaps (FX swaps in short) and cross-currency basis swaps (currency swaps in short) can serve similar purposes, though preferred by agents of different business models and cash flow durations.

The equation above conveys straightforward intuitions. Consider $(D^d - h^d)$, which represents the d-hedgers' *unhedged* pound exposure. We can also treat the quantity as if it is a pure speculative position on GBP/USD. This term increases in $(\mu_e + r^f - r^d)$, which is the (expected) excess return from a GBP/USD carry trade (borrowing dollars, exchanging for pounds in spot markets, then lending pounds). This excess return over the variance (scaled by the risk-aversion parameter) is the canonical mean-variance portfolio demand.

The d-hedgers' pure (unhedged) pound exposure $(D^d - h^d)$ increases when β_d, the regression coefficient of d-hedgers' endowment growth rates on the currency returns, decreases. Lower β_d makes the exchange rate E itself a better hedge against a future drop in d-hedgers' endowments, thus incentives hedgers to take on more the currency risk.

Hedged position h^d decreases in the CIP deviations. Recall that h^d represents the quantity of pounds d-hedgers are selling forward. As higher b translates to relatively lower forward pound price: selling pounds for dollar forward becomes less favorable, thus a smaller hedged position.

Similarly, for f-hedgers hedging against USD/GBP exchange risk, they will sell h^f units of dollar for h^f/F units of pounds forward. The resulting wealth change is

$$\begin{aligned} dW^f =& \frac{h^f}{F} + \left(D^f + \mathrm{d}D^f - h_t^f\right)\left(\frac{1}{E} + \mathrm{d}\left(\frac{1}{E}\right)\right) - \frac{D^f}{E} \\ =& \frac{h^f}{E}\left[\frac{E}{F} - 1 - E\mathrm{d}\left(\frac{1}{E}\right)\right] + \mathrm{d}\left(\frac{D^f}{E}\right) \\ =& \frac{h^f}{E}\left[\left(r^f - r^d + b + \mu_e - \sigma_e^2\right)\mathrm{d}t + \sigma_e dz_e\right] + \mathrm{d}\left(\frac{D^f}{E}\right). \end{aligned} \tag{C.4}$$

f-hedgers will choose h^f to maximize (C.1) for $j = f$ under their budget constraint (C.4), will lead to the following first-order condition:

$$\frac{1}{E}\left(r^f - r^d + b + \mu_e - \sigma_e^2\right) - A^f\left\{\left(\frac{1}{E}\right)^2 h^f\sigma_e^2 + \frac{\sigma_e}{E\mathrm{d}t}\mathbb{E}\left[\mathrm{d}z_e\mathrm{d}\left(\frac{D^f}{E}\right)\right]\right\} = 0.$$

From the equation above, we can solve for h^f:

$$\begin{aligned} h^f &= \frac{r^f - r^d + \mu_e - \sigma_e^2}{(A^f/E)\sigma_e^2} + \frac{b}{(A^f/E)\sigma_e^2} - \frac{E}{\sigma_e\mathrm{d}t}\mathbb{E}\left[\mathrm{d}z_e\mathrm{d}\left(\frac{D^f}{E}\right)\right] \\ &= \frac{r^f - r^d + \mu_e - \sigma_e^2}{A\sigma_e^2} + \frac{b}{A\sigma_e^2} - \frac{D^f}{\sigma_e^2\mathrm{d}t}\mathbb{E}\left[\frac{\mathrm{d}E}{E}\left(\frac{\mathrm{d}D^f}{D^f} - \frac{\mathrm{d}E}{E} + \frac{\mathrm{d}E}{E}\frac{\mathrm{d}E}{E} - \frac{\mathrm{d}D^f}{D^f}\frac{\mathrm{d}E}{E}\right)\right] \\ &= \frac{r^f - r^d + \mu_e - \sigma_e^2}{A\sigma_e^2} + \frac{b}{A\sigma_e^2} - \frac{D^f}{\sigma_e^2\mathrm{d}t}\mathbb{E}\left[\frac{\mathrm{d}E}{E}\frac{\mathrm{d}D^f}{D^f} - \sigma_e^2\mathrm{d}t\right] \\ &= \frac{\mu_e + r^f - r^d}{A\sigma_e^2} - \frac{1}{A} + \frac{b}{A\sigma_e^2} - \underbrace{\frac{\mathrm{Cov}\left[\mathrm{d}E/E,\, \mathrm{d}D^f/D^f\right]}{\mathrm{Var}\left[\mathrm{d}E/E\right]}}_{\beta_f} D^f + D^f. \end{aligned} \tag{C.5}$$

The f-hedgers optimal choice of h^f delivers similar intuitions as the d-hedgers'. Now that h^f represents the amount of dollars f-hedgers are selling forward for pounds, thus a long position on pounds, it increases in the (expected) GBP/USD risk premium, and decreases in β_f as defined above (a higher β_f means pounds do not offer protection against f-hedgers' endowment risk).

Based on equation (C.3) and (C.5), we can calculate the net demand for dollars in forward markets, in dollar terms. Since d-hedgers sell h^d units of pounds for dollars and f-hedgers sell h^f units of dollar for pounds, the net forward dollar demand is

$$h^d E - h^f = \underbrace{-\frac{2(\mu_e + r^f - r^d)}{A\sigma_e^2} + \frac{1}{A} + D^d E(1+\beta_d) - D^f(1-\beta_f)}_{\gamma_0} - \underbrace{\frac{2}{A\sigma_e^2}}_{\gamma>0} b.$$

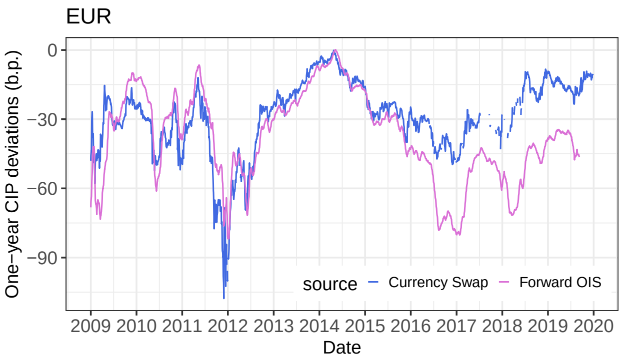
This expression agrees with the specification of hedgers' demand given in equation (5). Plugging in the assumptions that

$$\begin{aligned} D^d E &= \boldsymbol{\lambda}_d^\top \boldsymbol{x} + \lambda_{d,0}, \\ D^f &= \boldsymbol{\lambda}_f^\top \boldsymbol{x} + \lambda_{f,0}, \end{aligned}$$

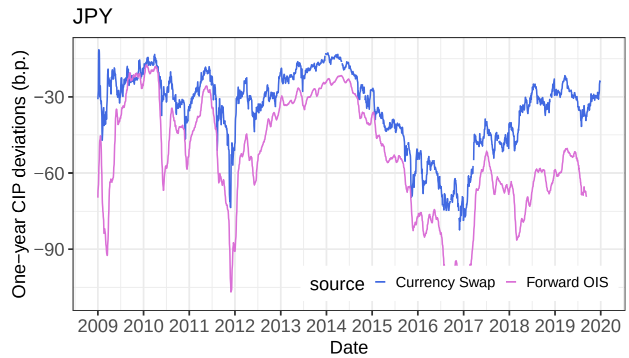
we have

$$\begin{aligned} h^d E - h^f = & \underbrace{\left[\boldsymbol{\lambda}_d^{(o)}(1+\beta_d) - \boldsymbol{\lambda}_f^{(o)}(1-\beta_f)\right]^\top}_{\boldsymbol{\beta}^\top} \boldsymbol{x}^{(o)} + \\ & \underbrace{\left[\boldsymbol{\lambda}_d^{(u)}(1+\beta_d) - \boldsymbol{\lambda}_f^{(u)}(1-\beta_f)\right]^\top \boldsymbol{x}^{(u)} + \lambda_{d,0}(1+\beta_d) - \lambda_{f,0}(1-\beta_f) + \frac{1}{A} - \frac{2(\mu_e + r^f - r^d)}{A\sigma_e^2}}_{\ell} - \gamma b, \end{aligned}$$

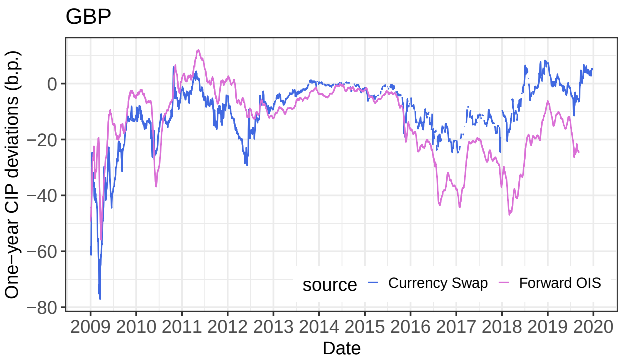
where $\boldsymbol{x}^{(o)}$ denotes observable components in $\boldsymbol{x}$ ($\boldsymbol{\lambda}_d^{(o)}$ and $\boldsymbol{\lambda}_f^{(o)}$ being loadings for the observable factors) and $\boldsymbol{x}^{(u)}$ denotes the unobservables ($\boldsymbol{\lambda}_d^{(u)}$ and $\boldsymbol{\lambda}_f^{(u)}$ being their loadings). Thus the net hedging demands have three parts: the linear combination of observable factors $\boldsymbol{\beta}^\top \boldsymbol{x}^{(o)}$, the latent unobservable demand ℓ, and the downward sloping response term $-\gamma b$.



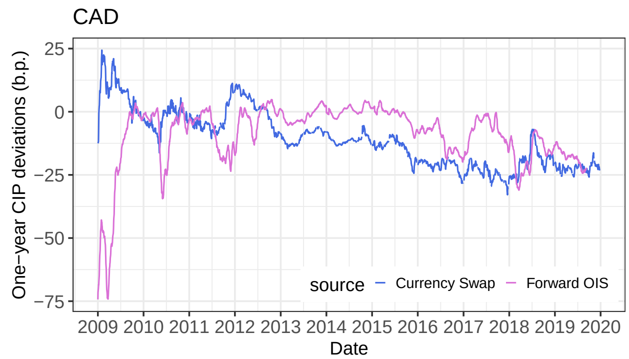
(A) EUR



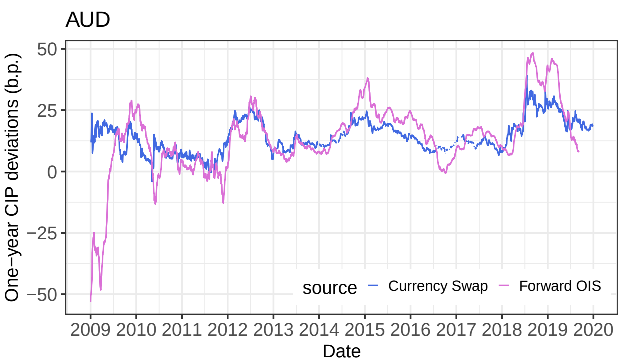
(B) JPY



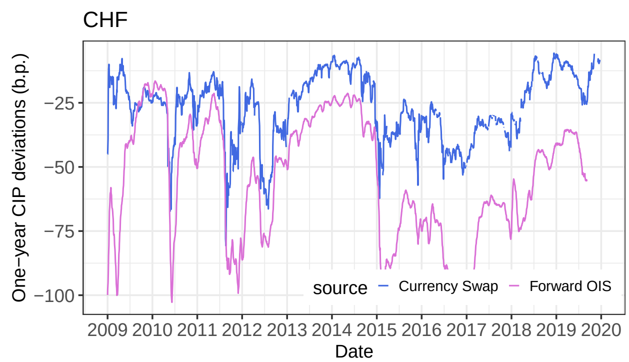
(C) GBP



(D) CAD



(E) AUD



(F) CHF

Figure A1: One-year CIP deviations for G6 currencies: currency swap rates and forward-OIS basis

D Additional tables and figures

Table A1: Dealer banks surveyed by foreign exchange committees, October 2004-April 2020

Australia and New Zealand Banking Group Limited	Bank of America Corporation
Bank of China	Bank of East Asia Limited
Bank of Montreal	Bank of New York Mellon Corporation
Bank of Nova Soctia	Barclays Plc
BNP Paribas SA	Canadian Imperial Bank of Commerce
China Bank of Communications	Citigroup Inc
Commerzbank AG	Commonwealth Bank of Australia
Crédit Agricole Corporate and Investment Bank	Credit Suisse Group AG
DBS bank Ltd	Deutsche Bank AG
Goldman Sachs Group Inc	Hang Seng Bank Limited
HSBC Holdings	Industrial and Commercial Bank of China
The ING Group	JP Morgan Chase & Co
Lloyds Banking Group Plc	Macquarie Bank Limited
Mizuho Bank Limited	Morgan Stanley
Mitsubishi UFJ Financial Group	National Australia Bank
National Bank of Canada	NatWest Group Plc
Nomura Holdings Inc	Oversea-Chinese Banking Corporation Limited
Resona Holdings Inc	Royal Bank of Canada
Shinsei Bank Limited	Skandinaviska Enskilda Banken AB
Société Générale SA	Standard Chartered Plc
State Street Corporation	Sumitomo Mitsui Financial Group Inc
Sumitomo Mitsui Trust Holdings Inc	Toronto-Dominion Bank
UBS AG	UniCredit SpA
United Overseas Bank Limited	Wells Fargo & Co
Westpac Banking Corporation	

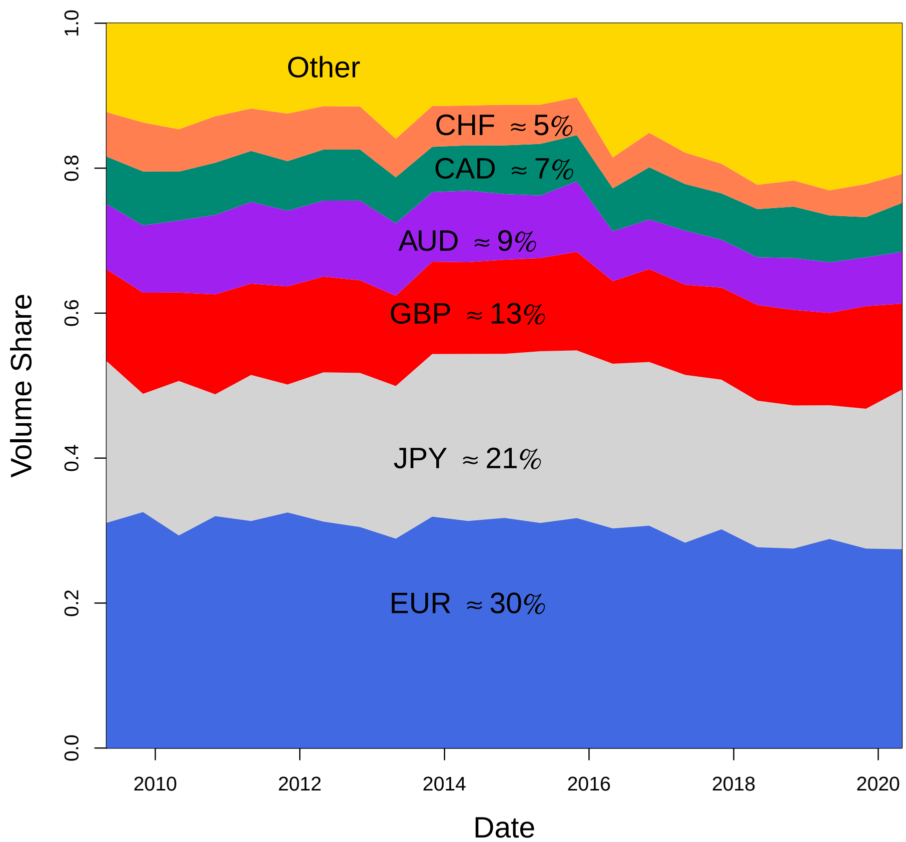


Figure A2: FX derivatives trading volume shares

Table A2: Predictive regressions: monthly returns of FX committee surveyed (FXS) dealer banks on one-year basis swap rates and placebo tests

This table presents results from the following linear regressions:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta|b_t| + \epsilon_{t+\tau},$$

for daily and monthly observations. The dependent variables are one-month-ahead value- or equal-weighted equity returns of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. Additional placebo tests use returns from five ETFs tracking the S&P500 index (SPY), the global financial sector (IXG), the US financial sector (IYF), US broker-dealers and securities exchanges (IAI), and US insurance companies (KIE). For monthly observations, five hedge fund index returns are also included: one global composite index from BarclaysHedge (BCH), four indices from Hedge Fund Research (HFR) tracking global composite, relative value arbitrage, global-macro, and macro-currency strategies. All returns are net ones in percentage, as well as annualized (divided by $\tau = 1/12$ as shown in the regression specification). The independent variable $|b_t|$ is the cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Sample periods begin from January 2009 and end at December 2019. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

Panel A: daily observations							
ret. (p.p.)	FXS (vw)	FXS (ew)	ETF-SPY (S&P500)	ETF-IXG (Gl. Fin.)	ETF-IYF (US Fin.)	ETF-IAI (US B&D)	ETF-KIE (US Insur.)
$\|b\|$ (b.p.)	2.17	2.03	0.41	1.85	1.18	1.30	1.11
	(0.82)	(0.81)	(0.41)	(0.78)	(0.70)	(0.80)	(0.73)
const.	−33.3	−31.0	7.8	−26.2	−8.4	−10.6	−4.7
	(15.2)	(15.1)	(8.9)	(14.7)	(13.5)	(17.0)	(14.5)
N obs.	2859	2859	2761	2761	2761	2761	2761
R^2-adj. (%)	3.5	3.4	0.3	2.8	1.4	1.6	1.1
Panel B: monthly observations							
ret. (p.p.)	FXS (vw)	FXS (ew)	ETF-SPY (S&P500)	ETF-IXG (Gl. Fin.)	ETF-IYF (US Fin.)	ETF-IAI (US B&D)	ETF-KIE (US Insur.)
$\|b\|$ (b.p.)	1.71	1.58	0.35	1.66	0.96	1.23	0.97
	(0.63)	(0.64)	(0.40)	(0.67)	(0.59)	(0.78)	(0.67)
const.	−23.4	−21.7	9.3	−22.1	−3.4	−8.6	−1.7
	(13.0)	(12.9)	(9.4)	(13.4)	(13.4)	(18.3)	(14.2)
N obs.	132	132	132	132	132	132	132
R^2-adj. (%)	1.8	1.7	−0.5	1.7	0.3	0.7	0.2
ret. (p.p.)			BCH (Gl. Com.)	HFR (Gl. Com.)	HFR (Re. Val.)	HFR (Macro)	HFR (Macro. Cur)
$\|b\|$ (b.p.)			0.13	0.08	0.09	−0.17	0.04
			(0.14)	(0.14)	(0.10)	(0.10)	(0.11)
const.			3.5	3.6	4.7	5.3	0.2
			(3.5)	(3.3)	(2.6)	(2.9)	(2.5)
N obs.			132	132	132	132	132
R^2-adj. (%)			−0.5	−0.6	−0.4	0.0	−0.7

Table A3: Predictive regressions: monthly returns of FX committee surveyed dealer banks on one-year basis swap rates adjusted by controls

This table presents results from the following linear regressions:

$$\frac{1}{\tau}\text{return}_{t+\tau} = \beta_0 + \beta\bar{b}_t + \phi \cdot \text{control}_t + \epsilon_{t+\tau}$$

for daily and monthly observations. The dependent variable is the one-month-ahead value-weighted equity return of 49 dealer banks surveyed by FX committees of New York, London, Tokyo, Toronto, Sydney, Singapore and Hong Kong. All returns are net ones in percentage, as well as annualized (divided by $\tau = 0.25$ as specified in the regression equation). The independent variable $\bar{b}_t$ is the cross-sectional average of absolute one-year basis swap rates for EUR, JPY, GBP, AUD, CAD, and CHF against the dollar. Control variables include the average smoothed earnings yield (E/P) and dividend yield (D/P) for the 49 dealer banks, the effective Fed fund rate (FFR), and the CBOE volatility index (VIX). Sample periods begin from January 2009 and end at December 2019. Numbers in parentheses are Newey-West standard errors under automatic bandwidth selection.

ret. (p.p.)	Daily observations			Monthly observations		
$\bar{b}$ (b.p.)	2.17	1.15	1.81	1.71	1.02	1.53
	(0.82)	(0.64)	(0.67)	(0.63)	(0.56)	(0.59)
E/P		14.7			10.8	
		(8.9)			(7.6)	
D/P			1.28			0.41
			(1.85)			(2.22)
FFR		5.03	0.26		3.17	0.20
		(6.99)	(6.27)		(6.59)	(6.92)
VIX		−0.44	1.95		−0.34	1.42
		(1.11)	(1.80)		(1.44)	(1.79)
const.	−33.3	−128.2	−65.1	−23.4	−93.4	−47.4
	(15.2)	(68.3)	(36.7)	(13.0)	(53.5)	(36.2)
N obs.	2859	2859	2859	132	132	132
R^2-adj. (%)	3.5	10.8	6.1	1.8	3.6	1.0