The PPP view of multihorizon currency risk premiums*

Mikhail Chernov† and Drew Creal‡

First draft: September 5, 2017
This draft: July 2, 2019

Abstract

Exposures of expected future depreciation rates to the current interest rate differential violate the UIP hypothesis in a distinctive pattern that is a non-monotonic function of horizon. Conversely, forward, or risk-adjusted expected depreciation rates are monotonic. We explain the two patterns jointly by incorporating the weak form of PPP, aka stationarity of the real exchange rate, into a joint model of the stochastic discount factor, the nominal exchange rate, inflation differential, domestic and foreign yield curves. Short-term departures from PPP generate the first pattern. The risk premiums for these departures generate the second pattern. Thus, the variance of the stochastic discount factor must be related to the real exchange rate deepening the exchange rate disconnect. We illustrate the challenge in the context of workhorse consumption-based models.

JEL Classification Codes: F31, F47, G12, G15.

Keywords: uncovered interest parity, purchasing power parity, cointegration, multiple horizons, affine term structure model, recursive preferences, habit-based preferences.

* Previously circulated as "Multihorizon currency returns and Purchasing Power Parity." We thank Michael Bauer, Magnus Dahlquist, Valentin Haddad, Lars Lochstoer, Rob Richmond, Rosen Valchev, Cynthia Wu, and Irina Zviadadze for comments on earlier drafts and participants in seminars at and conference sponsored by BIS, Boston Fed, Chicago Fed, Copenhagen, Dallas Fed, EDHEC, Erasmus, Northwestern, Oklahoma, SITE 2017, Tinbergen Institute, UT Dallas, UCLA Macro-Finance brown bag. The latest version is available at

https://sites.google.com/site/mbchernov/CC_PPP_latest.pdf.

† Anderson School of Management, UCLA, NBER, and CEPR; mikhail.chernov@anderson.ucla.edu.

‡ Department of Economics, University of Notre Dame; dcreal@nd.edu.

1 Introduction

The literature on foreign exchange (FX) rates has a strong interest in Uncovered Interest Parity (UIP) violations, that is, in documenting how their risk premiums vary with the state of the economy and what are the sources of this variation. More recently, this interest has expanded to multiple periods. One piece of evidence is that exposure of the forecasted depreciation rate to the respective interest rate differential (IRD) has a puzzling non-monotonic pattern, as a function of the forecast horizon (Bacchetta and van Wincoop, 2010; Engel, 2016; Valchev, 2016). In contrast, as we show in this paper, the same exposure of foreign forward rates is monotonic. The two pieces of evidence essentially reflect expectations of the same object – future depreciation rates – but under different probabilities, actual versus risk-adjusted, respectively.

In this paper we ask which features of the data generating process could account for the two types of patterns simultaneously. We conclude that incorporating weak, or "long-run", purchasing power parity (PPP, hereafter) into a joint model of exchange rates and bond prices goes a long way towards this goal. PPP posits that the real exchange rate (RER) is stationary, or, equivalently, that the nominal exchange rate between two countries and their respective price levels are cointegrated. The error correction component of this cointegrating relationship and the associated risk premium are the driving forces behind the empirical success of our model.

The error correction term characterizes how state variables adjust in response to a shock to the RER in order to restore the long-run relationship between these variables. The speed of this adjustment, which we denote by a parameter alpha, does not matter in the long run but will affect relationships between the state variables at intermediate horizons. Alphas are responsible for the documented shapes of exchange rate forecasts. Thus, any economic theory that is trying to explain the currency predictability patterns should be focusing on generating endogenous speed of error correction.

One might think that, because of the PPP effect, forward-based evidence should exhibit similar non-monotonic patterns across horizons. It does not. This is because forward rates reflect risk-adjusted expectations, and UIP holds under the risk-adjusted probability, a.k.a. covered interest parity (CIP). This means that the loading on the IRD is equal to one and that no other variable predicts the depreciation rate one step ahead: the risk-adjusted alpha controlling the speed of the depreciation rate's adjustment is equal to zero. This effect leads to a monotonic pattern in exposure to the IRD.

One implication of that result is that the variance of the stochastic discount factor (SDF) and, therefore, the maximum risk premium must increase with the decline in the RER (US consumption basket appreciation). This deepens the exchange rate disconnect, a broad set of evidence suggesting little to no relationship between exchange rates and the macroeconomy. That is because the risk premiums on financial assets are strongly related to business cycles while the RER is not (e.g., Fama and French, 1989, Meese and Rogoff, 1983).[1] Thus, one needs a structural model that simultaneously generates risk premiums that are related to the business cycle, and a RER that is not, all while the latter impacts the former. We explain the modeling challenge in the context of the workhorse open economy model of exchange rates with exogenous consumption and with recursive or habit-based preferences.[2]

We implement these ideas via a Gaussian no-arbitrage international term structure model. Such models are based on state variables that follow VAR-like dynamics. We complement such a VAR by the cointegrating relationship between the nominal exchange rate and the two price levels that is implied by PPP. The standard practice is to combine such a relationship with the VAR dynamics via a VECM representation. We reinterpret the VECM by introducing the companion form of a VAR and extending the original vector of state variables to include the new state, that is, the (stationary) RER. Including state variables that are cointegrated into the dynamics of the model is new to the literature on no-arbitrage term structure models. In order to be able to speak to the forward-rate evidence, we introduce an affine model of the SDF

First, we use this modeling framework to illustrate our main points in the context of the most simple setting. The state vector contains the nominal depreciation rate, inflation differential and the IRD. Second, we confirm these insights by estimating a realistic model of the joint behavior of bilateral depreciation rates and yield curves of the two respective countries. In this case, the state includes the nominal depreciation rate, the respective inflation rates, and nominal interest rates of two countries and of different maturities.

We estimate VARs in bilateral settings of the U.S.A. and one of the following four countries: the U.K., Canada, Germany/Euro, and Japan. In each case the model does a good job in capturing the joint dynamics of the macro variables and term structures of interest rates. In particular, the model replicates the cross-horizon regression patterns with respect to both actual and risk-adjusted expectations. The model also matches the long-horizon UIP patterns documented by Chinn and Meredith (2004). Further, we provide evidence that removing the PPP-based cointegrating relationship eliminates the model's ability to match the multi-horizon patterns.

Under the null of our model, the IRD-based regressions can be interpreted as projections of the nominal FX premium on the IRD. We compare the nominal FX premium implied by our model to the IRD-projected one. The two exhibit similar cyclical properties regardless of the horizon but the model-based premium is larger and often moves in the direction that is opposite of the projected one. This evidence suggests that UIP-based intuition about the FX premium behavior could be misleading and that there are other variables that could play an important role in the premium variation over time.

[1] To clarify, the model does not imply countercyclicality of the RER because the fitted RER follows the observed one by construction. Rather the model implies a movement in prices of risk that is unrelated to the basic macroeconomic aggregates. That is a challenge for existing theories.

[2] Engel (2016) rejects models with recursive preferences on the basis of the multi-horizon UIP evidence alone. As we discuss in the paper, all the considered models feature non-stationary RER. Any model with that feature does not need an SDF estimate to be rejected.

Related literature

The literature on currency risk premiums is extensive. In this brief review, we focus on papers that explore the role of the RER in this context. Our paper is most closely related to Dahlquist and Penasse (2016), who explore the PPP implications for UIP regressions. They impose PPP by iterating forward the relationship between nominal excess returns, real exchange rates, interest differentials, and inflation differential, a.k.a. the present value approach. They do not focus on the role of deviations from PPP at the short to intermediate horizons, neither do they study the interaction with yield curves. That prevents them from identifying the dynamics of the SDF, which we argue is challenging for existing equilibrium frameworks.

Engel (2016) offers the most forceful evidence of non-monotonicity of the UIP coefficients. He uses a VECM with the cointegrating relationship implied by PPP as a tool for constructing cross-horizon expectations of nominal depreciation rates, but there is no discussion of departures from PPP. As is the case with Dahlquist and Penasse (2016), his approach does not allow for identification of the SDF whose properties are central to our findings.

Jorda and Taylor (2012) use a similar VECM to motivate predictive regressions of nominal depreciation rates. Boudoukh, Richardson, and Whitelaw (2016) explore similar regressions. Balduzzi and Chiang (2017) explore the present value approach as a restriction that is helpful in increasing the power of tests of the UIP hypothesis. Relatedly, Barroso, Filipe, and Maio (2016) use the same restriction to compute variance decompositions of nominal exchange rates. Asness, Moskowitz, and Pedersen (2013); Menkhoff, Sarno, Schmeling, and Schrimpf (2017) use real exchange rates in the cross-section of currencies.

More broadly, our paper contributes to the literature on multi-horizon properties of financial assets, prominently summarized by Binsbergen and Koijen (2017). We demonstrate that the key modeling frameworks that use recursive or habit-based preferences have to confront a challenging set of evidence. Stepping out of these frameworks appears to be a promising avenue for future research on exchange rates.

2 Preliminary evidence

2.1 Review of regressions

Let S_t denote the exchange rate defined in terms of the number of dollars \$ per unit of foreign currency. If S_t increases, the U.S. dollar depreciates. Let ℓ_t and $\widehat{\ell}_t$ denote the one period U.S. and foreign interbank rates. We use lower case letters to denote the logarithm of a variable, i.e., $s_t = \log S_t$, and hats to denote a variable from a foreign country, i.e., $\widehat{\ell}_t$. (We reserve an asterisk, $*$, that is typically used to denote foreign variables in the international macro-finance literature, to denote the risk-adjusted probability and the associated parameters.)

We use $\Delta s_{t+1} = s_{t+1} - s_t$ to denote the one-period time-series difference operator, and $\Delta_c \ell_t = \ell_t - \widehat{\ell}_t$ to denote the cross-country difference operator. We study multiple countries, but we suppress asset-specific notation for simplicity.

The famous uncovered interest parity (UIP) regressions of Bilson (1981); Fama (1984); Tryon (1979) construct forecasts of next period depreciation rates, $E_t[\Delta s_{t+1}]$, on the basis of current IRDs. The recent literature, such as Engel (2016); Valchev (2016), focuses on forecasts of depreciation rates at longer horizons. That is, the authors document how $E_t[\Delta s_{t+n}]$ changes with horizon n as a function of $\Delta_c \ell_t$.

Specifically, these authors are motivated by the extension of UIP,

$$E_t(s_{t+1} - fs_t^0) = E_t(\Delta s_{t+1} - \Delta_c \ell_t) = 0$$

to multiple horizons. Here fs_t^{n-1} is the log price of the forward contract that pays $\$S_{t+n}/S_{t+n-1}$ per \$1 of notional at time $t+n$ in exchange for that forward price. As is well understood, the UIP hypothesis implicitly assumes constant prices of risk. Formally, the currency risk premium is computed in the levels of the variables, $E_t(e^{\Delta s_{t+1} - \Delta_c \ell_t}) = 1$ under UIP, and the expression in logs is obtained, under conditional log-normality, from this relationship, up to a convexity term:

$$\log E_t(e^{\Delta s_{t+1} - \Delta_c \ell_t}) = E_t(\Delta s_{t+1} - \Delta_c \ell_t) + Var_t(\Delta s_{t+1})/2. \tag{1}$$

In homoskedastic models, such as the one entertained in this paper, the convexity term is constant and, as a result, does not affect the slopes in UIP regressions, but formally speaking, it violates the strong form of the UIP hypothesis.

In order to extend the UIP hypothesis to Δs_{t+n}, consider the log risk premium on a currency forward contract

$$rps_t^n = \log E_t(e^{\Delta s_{t+n}}) - fs_t^{n-1}. \tag{2}$$

When $n = 1$ we recover equation (1). Again, up to convexity, the expression in (2) is the same studied by Engel (2016); Valchev (2016), who document that $cov(rps_t^n, \Delta_c \ell_t)$ switches sign with horizon n.

Thus, a different null hypothesis that accounts for the modern view of UIP could be: a theory of time-varying risk premiums that explains one-period UIP violations can also replicate the documented multi-horizon patterns. That view motivates us to refine the evidence and focus on each ingredient, $cov(E_t \Delta s_{t+n}, \Delta_c \ell_t)$ and $cov(fs_t^{n-1}, \Delta_c \ell_t)$ separately.

That is useful because the forward exchange rate can be viewed as the market's risk-adjusted expectation, denoted E_t^*, of the future depreciation rate. Denoting the $n-$period SDF by $\mathcal{M}_{t,t+n}$, we have

$$\begin{aligned}
rps_t^n &= \log E_t\left[e^{\Delta s_{t+n}}\right] - \log\left(E_t\left[\mathcal{M}_{t,t+n} e^{\Delta s_{t+n}}\right] / E_t\left[\mathcal{M}_{t,t+n}\right]\right) && (3)\\
&= \log E_t\left[e^{\Delta s_{t+n}}\right] - \log E_t^*\left[e^{\Delta s_{t+n}}\right] = E_t\left[\Delta s_{t+n}\right] - E_t^*\left[\Delta s_{t+n}\right] + \text{convexity} && (4)\\
&= -cov_t\left[\log \mathcal{M}_{t,t+n}, \Delta s_{t+n}\right] + \text{convexity}, && (5)
\end{aligned}$$

where lines (4) and (5) are two equivalent representations of the main equation (3). If $\mathcal{M}$ does not vary over time, the expression above is equal to zero, ignoring convexity, and UIP holds for horizon n. Thus we can characterize how two forecasts of depreciation rates, one under the true and the other one under the risk-adjusted probabilities, co-move with the IRD at different horizons. The advantage of this approach is that the evidence helps us to identify the properties of the SDF.

Traditionally, empirical analysis of exchange rates is implemented via regressions or VARs of excess returns on depreciation rates and potentially some other variables. Such analysis implies expectations of excess currency returns, a.k.a. currency risk premiums. This approach does not allow one to say which combinations of risks are being compensated for and how if the SDF is unknown.

For instance, in the case of $n = 1$, depreciation rates, Δs_{t+1}, could be driven by innovations in interest rates, inflation, and long-term interest rates. The attribution of risk premiums to these sources of risk can be achieved by decomposing expected excess depreciation rates, $E_t \left[\Delta s_{t+1} - \Delta_c \ell_t\right]$ (equation (1)), via covariances of the SDF with the respective innovations (equation (5)). Furthermore, one cannot model the behavior of asset prices, such as currency forward rates without considering the SDF. Thus, the regression/VAR-based evidence does not identify uniquely an economic mechanism that is responsible for observed risk premiums.

2.2 Data

We work with monthly data from the U.S., U.K., Canada, Germany/Eurozone, and Japan from January 1983 to December 2015 making for $T = 396$ observations per country. Nominal exchange rates are from the Federal Reserve Bank of St. Louis. Prior to the introduction of the Euro, we use the German Deutschemark and splice these series together beginning in 1999. Following a long tradition in the literature, interbank rate differentials $\Delta_c \ell_t$ are constructed from forward exchange rates, obtained from Datastream.[3] We obtained daily interbank rate and exchange rate data and take the last business day of each month.

To analyze the risk-adjusted forecasts, we need forward exchange rates at longer maturities. They are constructed from government bond yields, y_t^n, via $fy_t^{n-1} = ny_t^n - (n-1)y_t^{n-1}$. U.S. government yields are downloaded from the Federal Reserve and are constructed by Gurkaynak, Sack, and Wright (2007). All foreign government zero-coupon yields are downloaded from their respective central banks (Bank of England, Bundesbank, Bank of Canada and the Bank of Japan). All government yields have maturities of $12, 24, 36, 48, 60, 72, 84, 96, 108, 120$ months. Because the available maturities are annual, we can only run regressions on average annual forward rates. The quality, frequency, and available maturities of the government bond data dictate the choice of countries in our sample.

[3]Our analysis does not require the values of foreign rates, $\widehat{\ell}_t$. That allows us to avoid addressing the important analysis of CIP violations in Du, Tepper, and Verdelhan (2016).

2.3 Results

We implement UIP forecasting regressions of monthly changes in the depreciation rate on the IRD of the corresponding country

$$s_{t+n} - s_{t+n-1} = \gamma_0^n + \gamma^n \Delta_c \ell_t + u_{t+n}, \qquad n = 1, 2, \ldots, 120. \tag{6}$$

It is common to implement these regressions with fixed effects (γ^n is the same across countries). We report the average of country-specific γ^n because it is easier to construct a comparable measure from a model.[4] This regression departs from standard UIP regressions by using the depreciation rate as the left-hand side variable. Usually the left-hand side variable is the excess log return on a currency trade, that is, the one-period depreciation rate minus the IRD. UIP would predict $\gamma^1 = 1$ for our setup. A standard result in international finance is the 'UIP puzzle' which finds statistically significant negative estimated values of γ^1.

The blue lines of Figure 1 report the regression coefficients γ^n. They start below zero at a horizon of one month. They change sign and become positive at horizons of 3 to 8 years, before converging back towards zero. This evidence is consistent with the numbers presented in Engel (2016); Valchev (2016) and is viewed as a puzzle because it contradicts mainstream theories of exchange rates.

We can measure how the risk-adjusted expectation $E_t^* [\Delta s_{t+n}]$ is related to $\Delta_c \ell_t$ from a contemporaneous regression of forward exchange rates on the IRD. Given annual maturities of yields, our data on forward exchange rates measures the expected average annual change instead of monthly changes. The regression is

$$E_t^* [(s_{t+n} - s_{t+n-12})/12] \quad = \quad \gamma_0^{*n} + \gamma^{*n} \Delta_c \ell_t + u_t^{*n}, \qquad n = 12, 24, \ldots, 120, \tag{7}$$

where, again, we report the average value of country-specific γ^{*n}. At the one-month horizon, the forward exchange rate equals the IRD by no arbitrage. Consequently, UIP holds under the risk-adjusted probability, or, equivalently, CIP holds.

The red lines of Figure 1 report the regression coefficients γ^{*n}. In contrast to γ^n, they start positive near a value of one as expected from the CIP condition, decline monotonically, and never change sign. The presented evidence deepens the puzzle of the γ^n pattern.

2.4 Interpretation of the evidence

The regressions discussed above implicitly focus on the joint dynamics of the (log) depreciation rate Δs_t and the IRD $\Delta_c \ell_t$. If we focus our attention on simple models such as a

[4] In our dataset, the fixed-effect common γ^n and the reported average are similar. The comparison is available upon request.

vector autoregression of order one, then it is mathematically impossible to generate the documented non-monotonic pattern in the UIP regression coefficients if the joint dynamics of the two variables is not affected by anything else. Indeed, a vector autoregression (VAR) of order one would imply that regression coefficients are proportional to the powers of IRD's persistence – a monotonic pattern.

Appendix A discusses how, in a simple VAR model, one needs at least one more stationary variable that possesses the following properties in order to generate the observed patterns. First, this variable should either forecast Δs_{t+1} or $\Delta_c \ell_{t+1}$, or both. Second, the variable must be forecastable by $\Delta_c \ell_t$. These requirements are intuitive: one needs an extra variable forecasting the depreciation rate to break the monotonic pattern implying the first condition. However, the first condition on its own does not help at multiple horizons if the second one does not hold.

Third, the monotonic pattern in the risk-adjusted regression coefficients suggests that forecasting Δs_t is key. This is because the CIP condition,

$$\Delta_c \ell_t = f s_t^0 = \log E_t^* \left[\exp \left(\Delta s_{t+1} \right) \right],$$

implies that no variable, other than the interest rate differential $\Delta_c \ell_t$, forecasts Δs_{t+1}.[5] The difference between the actual and risk-adjusted worlds would be responsible for the difference in the patterns of regression coefficients.

In this paper we argue that the RER is a variable that satisfies these requirements. The cointegrating relationship between the nominal exchange rate and (log) price level differential implied by the stationarity of the RER, i.e. long-term PPP, guarantees that the first and second properties hold. Risk-adjustment takes care of the rest. In the following, we explicitly show how it works. While we cannot prove that there are no other variables that could satisfy the aforementioned conditions, we argue that none of the variables heretofore explored in the literature satisfy these requirements.

3 A simple model

The purpose of this section is to illustrate how the documented regression patterns can be replicated when the RER serves as a variable that co-moves with the nominal depreciation rate and IRD.

3.1 Error correction representation

We reconcile both actual and risk-adjusted patterns of the regression coefficients by highlighting the role of the risk of intermediate-term deviations from Purchasing Power Parity

[5] Under heteroscedasticity, the variance of the depreciation rate could be another predictor of Δs_t, but it is not forecastable by $\Delta_c \ell_t$, so it does not satisfy the first property.

(PPP). Short-term PPP states that the RER is equal to one, or, in logs, $e_t \equiv s_t - p_t + \widehat{p}_t = 0$, where p_t denotes the (log) price level. Short-term PPP does not hold empirically but there is a strong, although not universal, opinion that PPP does hold over the long-term, that is, e_t is stationary. We assert long-term PPP and show how this helps in understanding the evidence presented in the previous section.

We present a simple model motivated by the specifications of Engel (2016); Dahlquist and Penasse (2016); Jorda and Taylor (2012) that allows us to explain how PPP connects to the evidence. We introduce a vector of non-stationary macro variables $m_t = (s_t, \Delta_c p_t)^\top$, where $\Delta_c p_t = p_t - \widehat{p}_t$. Further, we work with the following stationary variables: domestic and foreign inflation rates $\pi_t = \Delta p_t$ and $\widehat{\pi}_t = \Delta \widehat{p}_t$, and their cross-sectional difference $\Delta_c \pi_t = \pi_t - \widehat{\pi}_t$; the IRD $\Delta_c \ell_t$. Stack the state variables into a vector f_t: $f_t = (\Delta m_t^\top, \Delta_c \ell_t)^\top$. RER $e_t = \beta_m^\top m_t$ is stationary, that is, the macro variables m_t are cointegrated with cointegrating vector $\beta_m^\top = (1, -1)$.

Ignoring means (assuming all variables have a zero mean), the state is assumed to follow a vector error correction model (VECM):

$$f_t = \Phi_f f_{t-1} + \alpha_f e_{t-1} + \Sigma_f \varepsilon_t.$$

Errors in this model are deviations from the cointegrating relation $e_t = 0$ (long-term PPP). They set in motion changes in f_t that correct the errors. The vector $\alpha_f = (\alpha_s, \alpha_\pi, \alpha_\ell)^\top$ controls the speed of this error correction.

To simplify the setup, assume that

$$\Phi_f = \begin{pmatrix} 0 & 0 & \phi_{s\ell} \\ 0 & \phi_\pi & 0 \\ 0 & 0 & \phi_\ell \end{pmatrix}, \quad \Sigma_f = \begin{pmatrix} \sigma_s & 0 & 0 \\ 0 & \sigma_\pi & 0 \\ 0 & 0 & \sigma_\ell \end{pmatrix}.$$

The coefficient $\phi_{s\ell}$ is related to the UIP regression. The RER follows

$$\begin{aligned} e_t &= \beta_m^\top m_t = e_{t-1} + \Delta s_t - \Delta_c \pi_t \\ &= -\phi_\pi \Delta_c \pi_{t-1} + \phi_{s\ell} \Delta_c \ell_{t-1} + (1 + \alpha_s - \alpha_\pi) e_{t-1} + \sigma_s \varepsilon_{st} - \sigma_\pi \varepsilon_{\pi t}. \end{aligned}$$

As a result we can re-write the VECM as a (restricted) VAR by creating a new state vector $x_t = (f_t^\top, e_t)^\top$. The dynamics of x_t in companion form are:

$$x_t = \Phi_x x_{t-1} + \Sigma_x \varepsilon_t \tag{8}$$

with

$$\Phi_x = \begin{pmatrix} 0 & 0 & \phi_{s\ell} & \alpha_s \\ 0 & \phi_\pi & 0 & \alpha_\pi \\ 0 & 0 & \phi_\ell & \alpha_\ell \\ 0 & -\phi_\pi & \phi_{s\ell} & 1 + \alpha_s - \alpha_\pi \end{pmatrix}, \quad \Sigma_x = \begin{pmatrix} \sigma_s & 0 & 0 \\ 0 & \sigma_\pi & 0 \\ 0 & 0 & \sigma_\ell \\ \sigma_s & -\sigma_\pi & 0 \end{pmatrix}.$$

One obvious advantage of this companion form is that valuation of bonds is straightforward in the no-arbitrage framework.

Further, if the RER is stationary, the companion form makes it clear that at least one of the α_f's must be non-zero. Therefore, e_t must forecast at least one element of f_t. This is a manifestation of the first property highlighted in section 2.4. In a univariate regression setting, Dahlquist and Penasse (2016) emphasize that e_t is helpful in forecasting Δs_{t+1}, that is, $\alpha_s \neq 0$. The second property holds as well: the IRD $\Delta_c \ell_t$ forecasts e_{t+1} as long as it forecasts Δs_{t+1}. This is because of the PPP-implied restriction $\phi_{e\ell} \equiv \Phi_{x43} = \Phi_{x13} \equiv \phi_{s\ell}$.

Finally, the VAR representation implies that the relationship between horizon n and the forecast $E_t [\Delta s_{t+n}]$ is controlled by exponents of the matrix Φ_x, which is affected by the properties of α_f. Indeed,

$$E_t [\Delta s_{t+n}] = \mathrm{e}_1^\top \Phi_x^n x_t, \quad \mathrm{e}_1^\top = (1, 0, 0, 0).$$

In general, it is difficult to obtain tractable closed-form expressions for long horizons n. We can do so for horizons $n = 1, 2, 3$ in the case of our simple model:

$$\begin{aligned}
E_t [\Delta s_{t+1}] &= \phi_{s\ell} \Delta_c \ell_t + \alpha_s e_t, && (9)\\
E_t [\Delta s_{t+2}] &= -\phi_\pi \alpha_s \Delta_c \pi_t + (\phi_{s\ell}\phi_\ell + \phi_{e\ell}\alpha_s)\Delta_c \ell_t + (\phi_{s\ell}\alpha_\ell + \alpha_s \alpha_e) e_t, && (10)\\
E_t [\Delta s_{t+3}] &= -\phi_\pi(\alpha_s^2 + \alpha_s\alpha_e + \phi_{s\ell}\alpha_\ell)\Delta_c \pi_t + [\phi_{e\ell}\alpha_s(\alpha_e + \alpha_s) + \phi_{s\ell}(\phi_\ell^2 + \phi_{e\ell}\alpha_\ell)]\Delta_c \ell_t \\
&\quad + [\phi_{s\ell}\alpha_\ell(\alpha_e + \phi_\ell) + \alpha_s(\alpha_e^2 - \phi_\pi \alpha_\pi + \phi_{e\ell}\alpha_\ell)] e_t, && (11)
\end{aligned}$$

where $\alpha_e \equiv 1 + \alpha_s - \alpha_\pi$, and we used $\phi_{e\ell}$, which is equal to $\phi_{s\ell}$ under PPP, to emphasize the hypothetical case of $\phi_{e\ell} = 0$. The expression in (9) highlights "the missing premium" of Dahlquist and Penasse (2016). The expressions in (10), (11) make the role of α_f for forecasting obvious. Even if only $\alpha_s \neq 0$, it impacts the forecasting ability of all elements of x_t.

As the horizon increases, the loadings of $E_t [\Delta s_{t+n}]$ on $\Delta_c \ell_t$ can be written as the sum of two terms that are controlled by the forecasting parameters $\phi_{s\ell}$ and α_f. We highlight these here for $n = 3$ as

$$\begin{aligned}
\text{term 1} &= \phi_{s\ell}(\phi_\ell^2 + \phi_{e\ell}\alpha_\ell) \\
\text{term 2} &= \phi_{e\ell}\alpha_s(\alpha_e + \alpha_s).
\end{aligned}$$

The first term contains $\phi_{s\ell}$ and it multiplies powers of the IRD autocorrelation coefficient ϕ_ℓ^2, which becomes ϕ_ℓ^{n-1} at longer horizons. This term induces a slow monotonic decay in the covariances as the horizon increases and it is the dominant component of the UIP regression coefficients γ^n, especially at short horizons. If the RER were not present in the model ($\alpha_f = 0$), or if $\phi_{e\ell} = 0$ then the cross auto-covariance between the depreciation rate and the IRD would simply decay monotonically because it is influenced only by the product $\phi_{s\ell}\phi_\ell^{n-1}$ as the horizon increases. These observations are consistent with the properties outlined in section 2.4.

In order to illustrate these relationships quantitatively, we estimate the VECM in (8) using the U.S. and the U.K. data. In the spirit of the previous section, we report the model-implied coefficients γ^n. The results are presented in the left panel of Figure 2.

We consider several scenarios to emphasize the role of α_f : (i) all elements of α_f are equal to zero; (ii) only one of the elements of α_f is not equal to zero; (iii) all the elements of α_f are free. The first case corresponds to the regular VAR for the state f_t. It implies the standard pattern of monotonically increasing coefficients that approach zero at long horizons. The second case when $\alpha_\pi \neq 0$ happens to be almost identical. The coefficients γ^n cross zero at long horizons suggesting a potential hump at $n > 120$ months when $\alpha_\ell \neq 0$. Finally, when $\alpha_s \neq 0$ and in the third case, we observe the pattern that is qualitatively consistent with Figure 1.

3.2 Risk adjustment

The physical distribution of the state vector x_t implied together with the SDF yield the risk-adjusted distribution of x_t via $p^*(x_{t+1}|x_t)/p(x_{t+1}|x_t) = \mathcal{M}_{t,t+1}/E_t[\mathcal{M}_{t,t+1}]$. Our model is too simple to perform a formal risk-adjustment because we do not have an explicit specification of the reference interest rate ℓ_t. Therefore, we follow a storied finance tradition and assume $p^*(x_{t+1}|x_t)$ directly. The "volatility" matrix Σ_f is unchanged. The persistence matrix Φ_f^* could be different from Φ_f, including the zero elements becoming non-zero.

For the purposes of this discussion we simplify and assume the following form:

$$\Phi_x^* = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \phi_\pi^* & 0 & \alpha_\pi^* \\ 0 & 0 & \phi_\ell^* & \alpha_\ell^* \\ 0 & -\phi_\pi^* & 1 & 1-\alpha_\pi^* \end{pmatrix}.$$

The first row is dictated by the fact that CIP must hold. The second and third rows are assumed. The last row is implied by the first three from the definition of the RER.

The CIP-imposed restrictions that $\phi_{s\ell}^* = 1$ and $\alpha_s^* = 0$ immediately suggest that the risk-adjusted pattern could have different properties. Indeed, the risk-adjusted counterparts of terms 1 and 2 in equation (11) for a forecast horizon $n = 3$ are:

$$\begin{aligned} \text{term } 1^* &= \phi_\ell^{*2} + \alpha_\ell^* \\ \text{term } 2^* &= 0. \end{aligned}$$

If α_ℓ^* is sufficiently close to zero, we obtain a monotonic pattern in the regression coefficients, γ^{*n} that starts at a value of one at horizon $n = 1$ due to CIP.

One needs to use prices of market instruments, e.g., bonds, to estimate the risk-adjusted parameters. We are not going to do that in this section. Instead, we simply assume that the state variables are more persistent under the risk-adjusted probability. Thus, we set $\phi_\ell^* = 0.99$ ($\phi_\ell = 0.97$) and $\phi_\pi^* = 0.5$ ($\phi_\pi = 0.27$). We further consider two scenarios with either $\alpha_\pi^* = \alpha_\ell^* = 0$, or $\alpha_\pi^* = \alpha_\pi$ and $\alpha_\ell^* = \alpha_\ell$. We set $\alpha_s^* = 0$ in both scenarios because of CIP.

The right panel of Figure 2 displays the results. As a benchmark, the red line with crosses shows the actual pattern of γ^n corresponding to the full VECM model from the left panel. The green line with asterisks corresponds to the case when $\alpha_\pi^* = \alpha_\pi$ and $\alpha_\ell^* = \alpha_\ell$. This line is monotonic but its slope appears to be too small compared to the evidence in Figure 1. Most importantly, the values of γ^{*n} for large n are much higher than the corresponding γ^n. The black line with squares corresponds to $\alpha_\pi^* = \alpha_\ell^* = 0$. In this case the pattern is qualitatively much closer to the empirical one.

4 A realistic model

We have presented multihorizon empirical patterns of coefficients that relate actual and risk-adjusted expectations of future depreciation rates to the current IRD. We illustrated, using a simple model, how these patterns can be captured in one framework by incorporating the RER that converges to PPP in the long run and currency risk premiums. In this section we verify that this intuition actually holds in the data by developing an international no-arbitrage term structure model of nominal yields together with inflation rates, and nominal and real exchange rates.

We follow a plan that is similar to the presentation of the simple model in section 3. We start with a generic state f_t that controls the dynamics of the state variables and follows an error correction model. We show how it is related to macro variables and, after properly adjusting for risk premiums, to domestic and foreign bond prices. Then we present a specific choice of the state f_t whose elements are easily interpretable. To the best of our knowledge, the VECM structure for the factors and its companion form are new to the literature on no-arbitrage term structure models. The literature on international no-arbitrage term structure models does not incorporate the real exchange rate as a factor.

4.1 State dynamics

We specify the dynamics of the state f_t as a Gaussian VECM given by

$$f_t = \mu_f + \Phi_f f_{t-1} + \Pi_f f_{t-1}^L + \Sigma_f \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1) \tag{12}$$

where f_t^L denotes the factors in levels. The factors f_t are stationary while the levels f_t^L are unit-root non-stationary. This implies the existence of cointegration and that the matrix of coefficients Π_f has reduced rank; see Engle and Granger (1987). It can be factored as $\Pi_f = \alpha_f \beta_f^\top$ where β_f is the matrix of cointegrating vectors. The matrix α_f contains the speed of adjustment parameters that determine how fast the system converges back to its long-run equilibrium. Our model of cointegration is an example of an error correction representation; see, e.g. equation [19.1.42] in Hamilton (1994). Our representation differs from the standard approach in the econometrics literature in two ways. We define f_t to include only $I(0)$ variables rather than a mixture of $I(1)$ and $I(0)$ variables and we define the matrix of cointegrating vectors β_f to include only linear combinations of non-stationary variables; see Appendix B for more discussion.

4.1.1 Macro variables

We model the depreciation rate and the inflation rate differential as a linear function of the state given by

$$\Delta s_t = \delta_{s,0} + \delta_{s,f}^\top f_t \tag{13}$$

$$\Delta_c \pi_t = \delta_{\pi,0} + \delta_{\pi,f}^\top f_t. \tag{14}$$

For convenience, we stack the nominal exchange rates and price level differentials into a vector $m_t = (s_t \;\Delta_c p_t)^\top$ and write their first differences Δm_t as a function of the factors as

$$\Delta m_t = \delta_{m,0} + \delta_{m,f} f_t. \tag{15}$$

The initial value $m_0 = (s_0 \;\Delta_c p_0)^\top$ is assumed to be known. The log RER between the U.S. and foreign country is defined as

$$e_t \equiv s_t - \Delta_c p_t \equiv \beta_m^\top m_t, \tag{16}$$

where $\beta_m^\top = (1 \;\; -1)$.

4.1.2 Companion form of state dynamics

Given the relationship between the macroeconomic variables m_t and the state variables f_t, the dynamics of the real exchange rates e_t are pinned down by the dynamics of the factors f_t in (12). To see this, we write real exchange rates in terms of the factors

$$e_t = \beta_m^\top m_t = e_{t-1} + \beta_m^\top (\delta_{m,0} + \delta_{m,f} f_t)$$

and substitute in f_t from (12) to find the dynamics of real exchange rates

$$e_t = \beta_{f,0} + \beta_f^\top \mu_f + \beta_f^\top \Phi_f f_{t-1} + \left(1 + \beta_f^\top \alpha_f\right) e_{t-1} + \beta_f^\top \Sigma_f \varepsilon_t \tag{17}$$

where $\beta_f^\top = \beta_m^\top \delta_{m,f}$ and $\beta_{f,0} = \beta_m^\top \delta_{m,0}$.

Combining (12) and (17), we define the state vector $x_t = \left(f_t^\top \; e_t^\top\right)^\top$ and write the VECM in companion VAR form

$$x_t = \mu_x + \Phi_x x_{t-1} + \Sigma_x \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1), \tag{18}$$

where the vectors and matrices are defined as

$$\mu_x = \begin{pmatrix} \mu_f \\ \beta_{f,0} + \beta_f^\top \mu_f \end{pmatrix} \quad \Phi_x = \begin{pmatrix} \Phi_f & \alpha_f \\ \beta_f^\top \Phi_f & 1 + \beta_f^\top \alpha_f \end{pmatrix} \quad \Sigma_x = \begin{pmatrix} \Sigma_f \\ \beta_f^\top \Sigma_f \end{pmatrix}.$$

The companion form for x_t makes immediately clear that if $\alpha_f = 0$ so that there is no cointegration then the RER e_t must be non-stationary. This is because Φ_x reduces to a lower block-triangular matrix whose lower right block is simply equal to one when $\alpha_f = 0$. The matrix Φ_x will have (at least) one eigenvalue equal to one. Conversely, if e_t is stationary, then $\alpha_f \neq 0$ and the real exchange rate must forecast at least one of the variables in the system: future depreciation rates, inflation rate differentials, or interest rates.

Most theories in the international macroeconomics literature result in stationary real exchange rates. In the sequel, we confirm this by testing the null hypothesis of $\alpha_f = 0$. A natural question to address next is which other variable the RER forecasts. This point is similar to Cochrane (2008), where the price-to-dividend ratio represents the cointegrating relationship. If it is stationary, then it must forecast either returns or dividend growth.

4.2 Yields

It is standard practice in the literature to run the UIP regressions using interbank rates as the one month IRD. While researchers frequently associate these rates with Libor, this interpretation is problematic prior to Libor's inception in 1986 and in the wake of the financial crisis of 2008 (Du, Tepper, and Verdelhan, 2016). We describe how we address these issues in the implementation section. We refer to the relevant U.S. interbank rate as U.S. Libor, for brevity. We use the U.S. Libor rate as the reference discount rate so that we could speak to the UIP regressions directly. Subsequently, we derive all other bond prices relative to this curve.

4.2.1 The stochastic discount factor and risk-adjusted distribution

We model the dynamics of the log SDF denominated in terms of the U.S. Libor rate as

$$\log \mathcal{M}_{t,t+1} = -\delta_{\ell,0} - \delta_{\ell,x}^\top x_t - \frac{1}{2}\lambda_t^\top \lambda_t - \lambda_t^\top \varepsilon_{t+1} \tag{19}$$

with market prices of risk

$$\lambda_t = \Sigma_f^{-1}\left(\lambda_0 + \lambda_f f_t + \lambda_e e_t\right). \tag{20}$$

See Appendix C. Thus, if λ_e is significantly different from zero, the variance of the SDF, a.k.a. the maximum risk premium, is affected by the real exchange rate. The same observation applies to the real SDF unless the variance of inflation is such that this effect cancels out.

The physical distribution of the state vector x_t implied by (18) together with the SDF (19) yield the risk-adjusted distribution of x_t

$$x_t = \mu_x^* + \Phi_x^* x_{t-1} + \Sigma_x \varepsilon_t.$$

The matrices of parameters under the risk-adjusted probability share a similar form as above

$$\mu_x^* = \begin{pmatrix} \mu_f^* \\ \beta_{f,0} + \beta_f^{*\top} \mu_f^* \end{pmatrix} \qquad \Phi_x^* = \begin{pmatrix} \Phi_f^* & \alpha_f^* \\ \beta_f^{*\top} \Phi_f^* & 1 + \beta_f^{*\top} \alpha_f^* \end{pmatrix} \qquad \Sigma_x = \begin{pmatrix} \Sigma_f \\ \beta_f^{*\top} \Sigma_f \end{pmatrix}$$

where

$$\mu_f^* = \mu_f - \lambda_0 \qquad \Phi_f^* = \Phi_f - \lambda_f \qquad \alpha_f^* = \alpha_f - \lambda_e$$

The speed of adjustment parameters α_f may carry a risk premium.

In our setting, the matrices containing the cointegrating vectors $\beta_f = \beta_f^*$ are the same across probability measures, which gives the real exchange rate the same definition. It is possible to write down a more general model where there may exist cointegrating relationships across yields, price levels, exchange rates, and other macroeconomic variables. A researcher could then estimate $\left(\beta_f, \beta_f^*\right)$ and test for the presence of cointegrating relationships across series and across countries. We leave this extension to future research and focus on the setting where the only cointegrating relationships in the model are those defined by the real exchange rates in (37).

4.2.2 Libor-related rates

The prices of hypothetical zero-coupon U.S. and foreign Libor bonds with maturity n are given by the standard pricing condition

$$L_t^n = E_t^* \left[e^{-\delta_{\ell,0} - \delta_{\ell,x}^\top x_t} L_{t+1}^{n-1} \right]. \tag{21}$$

$$\widehat{L}_t^n = E_t^* \left[e^{-\delta_{\ell,0} - \delta_{\ell,x}^\top x_t} \frac{S_{t+1}}{S_t} \widehat{L}_{t+1}^{n-1} \right]. \tag{22}$$

U.S. and foreign yields $\ell_t^n = -n^{-1} \log L_t^n$ and $\widehat{\ell}_t^n = -n^{-1} \log \widehat{L}_t^n$ of all maturities n are linear functions of the factors

$$\ell_t^n = a_n + b_{n,x}^\top x_t, \tag{23}$$

$$\widehat{\ell}_t^n = \widehat{a}_n + \widehat{b}_{n,x}^\top x_t. \tag{24}$$

Expressions for the bond loadings can be found in Appendix D. By writing the model in companion form, they have the same expressions as standard Gaussian ATSMs, see, e.g., Ang and Piazzesi (2003). We reserve notation without superscript for the one-period yield, $\ell_t \equiv \ell_t^1$.

4.2.3 Government yields

It is well known that there exists a spread between short-term interbank rates (Libor) and and interest rates implicit in bonds issued by government institutions. At the one month horizon, this is the well-known Ted spread which is a popular way of measuring the credit quality of large financial institutions. The Ted spread also reflects a liquidity premium embedded in U.S. Treasuries.

To solve for bond prices, we use the results from Duffie and Singleton (1999) that imply the following prices for government bonds

$$Q_t^n = E_t^* \left[e^{-(\ell_t - c_t)} Q_{t+1}^{n-1} \right], \tag{25}$$

$$\widehat{Q}_t^n = E_t^* \left[e^{-(\ell_t - \widehat{c}_t)} \frac{S_{t+1}}{S_t} \widehat{Q}_{t+1}^{n-1} \right], \tag{26}$$

where c_t and $\widehat{c}_t$ are domestic and foreign credit/liquidity risk factors reflecting the product of risk-adjusted default probability and loss given default, and a liquidity component. We model these as a linear function of the state vector

$$c_t = \delta_{c,0} + \delta_{c,x}^\top x_t, \tag{27}$$

$$\widehat{c}_t = \widehat{\delta}_{c,0} + \widehat{\delta}_{c,x}^\top x_t. \tag{28}$$

Foreign and domestic government yields $y_t^n = -n^{-1} \log Q_t^n$ and $\widehat{y}_t^n = -n^{-1} \log \widehat{Q}_t^n$ are linear in the state variables

$$\begin{aligned} y_t^n &= d_n + h_{n,x}^\top x_t, \\ \widehat{y}_t^n &= \widehat{d}_n + \widehat{h}_{n,x}^\top x_t. \end{aligned}$$

with $y_t \equiv y_t^1$. Expressions for the bond loadings are in Appendix D.

The Ted spread is then measured by $c_t = \ell_t - y_t$ and with hats for its foreign counterpart. As is the case with interest rates themselves, the Ted spread could in theory become negative in our Gaussian model. In practice, the fitted values are positive. A final caveat is that, formally speaking, the SDF in (19) has to be adjusted to reflect an additional compensation for the combined default/liquidity risk. In practice, this risk premium cannot be identified well because of the rarity of defaults of banks on the Libor panel. As a result, we can only infer risk-adjusted default probabilities embedded in the Ted spread. For this reason, we simplify the notation and ignore the default component of the SDF.

4.3 Choice of state

The full state is $x_t = (f_t^\top, e_t)^\top$ as before. In this subsection, we describe a particular choice of the state vector f_t that is similar to the VAR tradition in macroeconomics. Specifically, we define the state vector as

$$f_t^\top \;=\; \left(\; \Delta s_t,\;\; \Delta_c\pi_t,\;\; \ell_t,\;\; y_t^{120,12},\;\; c_t,\;\; \Delta_c\ell_t,\;\; \Delta_c y_t^{120,12},\;\; \Delta_c\ell_t^{12,1} \;\right) \tag{29}$$

The factors are all observable a priori and, in addition to macro variables, include the domestic yield variables: the U.S. Libor rate ℓ_t, the U.S. government term spread $y_t^{120,12} = y_t^{120} - y_t^{12}$, the one month U.S. Ted spread c_t; and the variables capturing differences in yield curves across countries: the one-month Libor differential $\Delta_c\ell_t$, the differential in term spreads $\Delta_c y_t^{120,12} = y_t^{120,12} - \widehat{y}_t^{120,12}$, and the difference in (short) term spreads of the Libor curve $\Delta_c\ell_t^{12,1} = \ell_t^{12,1} - \widehat{\ell}_t^{12,1}$.

The large number of yield factors is due to the fact that we are modeling both domestic and foreign yield curves as well as the Libor differentials. Jiang, Krishanmurthy, and Lustig (2018), in independent work, also consider the relation of the differences between Treasury and Libor rates to currency risk premiums. In our notation, their factor, termed Treasury basis, is equal to $\widehat{c}_t - c_t$, or $\Delta_c y_t^n - \Delta_c\ell_t^n$ for longer maturities. See Du, Im, and Schreger (2018) for a similar measure (Treasury Premium). Finally, our choice of the state vector intentionally nests the simple model of section 3, where the state vector is $f_t^\top = (\Delta s_t,\; \Delta_c\pi_t,\; \Delta_c\ell_t)$ and yields of longer maturity are dropped from the model.

4.4 Identifying restrictions

We develop restrictions on the model that guarantee the elements of x_t have the interpretation we have selected. In this section, we briefly discuss some of these identifying restrictions. Appendix E contains the full details.

In our model, all the state variables in x_t are observable. The free parameters that govern the dynamics of the state, μ_x, Φ_x, Σ_x, are identifiable directly from the vector error correction model. These parameters therefore require no identifying restrictions. Restrictions are required on the factor loadings and the risk-adjusted parameters μ_x^*, and Φ_x^*.

Let e_j denote a unit vector with a one in location j and zeros in all other entries. The factor loadings and intercepts for the macroeconomic variables, Libor rate, and credit spread are restricted as follows:

$$\delta_{s,0} = 0, \quad \delta_{s,x} = \mathrm{e}_1, \tag{30}$$

$$\delta_{\pi,0} = 0, \quad \delta_{\pi,x} = \mathrm{e}_2, \tag{31}$$

$$\delta_{\ell,0} = 0, \quad \delta_{\ell,x} = \mathrm{e}_3, \tag{32}$$

$$\delta_{c,0} = 0, \quad \delta_{c,x} = \mathrm{e}_5, \tag{33}$$

Each of these restrictions results naturally from placing the observables $(\Delta s_t, \Delta_c \pi_t, \ell_t, c_t)$ in the state vector x_t. The rows of μ_x^* and Φ_x^* associated with these four variables all contain free parameters.

The IRD $\Delta_c \ell_t$ is also an element of the state vector in (29). Consequently, the risk-adjusted parameters must satisfy the following restrictions:

$$\mu_{x,1}^* = -\frac{1}{2} \mathrm{e}_1^\top \Sigma_x \Sigma_x^\top \mathrm{e}_1, \quad \mathrm{e}_1^\top \Phi_x^* = \mathrm{e}_6^\top, \tag{34}$$

This restriction can be viewed as an enforcement of the CIP condition. Indeed, equations (23) and (24) imply that for $n = 1$, the IRD is

$$\Delta_c \ell_t = -\delta_{s,0} - \delta_{s,x}^\top \mu_x^* - \frac{1}{2} \delta_{s,x}^\top \Sigma_x \Sigma_x^\top \delta_{s,x} - \delta_{s,x}^\top \Phi_x^* x_t.$$

See Appendix D. After imposing restriction (30), we see that (34) must hold in order for $\Delta_c \ell_t$ to be an entry of x_t. The restriction (34) forces the parameters in the first row of μ_x^* and Φ_x^* to be equal to either zero, one, or a deterministic function of other parameters of the model, e.g. the variance of the depreciation rate.

The remaining rows of μ_x^* and Φ_x^* are in general non-zero, but not all of the parameters in these rows are freely estimable. Instead, some rows of μ_x^* and Φ_x^* are deterministic non-linear functions of the parameters in other rows. Specifically, the three rows of μ_x^* and Φ_x^* associated with the term spreads in (29) are functions of parameters in other rows. Intuitively, an asset pricing equation (21) imposes internal consistency across yields of different maturities. No-arbitrage implies that yields of longer maturity are risk-adjusted forecasts of future short term interest rates, where forecasts are made using the model of the short rate ℓ_t. Therefore, the rows of μ_x^* and Φ_x^* associated with longer term yields are pinned down by this relationship.

Such restrictions make it challenging to parameterize the matrix Φ_x^* directly. The term structure literature solves this problem by parameterizing the matrix Φ_x^* in terms of a latent factor representation as in Joslin, Singleton, and Zhu (2011). We extend their results for vector autoregressions to vector error correction models.

While parameterizing the risk-adjusted parameters μ_x^* and Φ_x^* in terms of the latent factors makes estimation easier, the interpretation of the estimates under this rotation is challenging. Therefore, we use the latent factor parameterization to estimate the model but we report the more meaningful estimates of Φ_x^* implied by the observable parameterization.

4.5 Empirical approach

In this subsection, we describe the data that we use in addition to what is described in section 2, how the model is related to the data via the state-space representation, and which versions of our model we estimate.

While we refer to ℓ_t as U.S. Libor, we have to be careful with the data that we use to represent the U.S. interbank rate in different periods. Prior to 1986 we use the data from Engel (2016). We use U.S. Libor that was downloaded from the Federal Reserve Bank of St. Louis from 1986 to 2007 (similar to Engel's data during the corresponding period). Because forward rate transactions are fully collateralized, the market participants started using the overnight index swap (OIS) rate at the end of 2007 and the whole industry has switched to OIS by the end of 2008. We reflect this change, by using OIS rates as a measure of ℓ_t starting in 2009, and by using a weighted average of Libor and OIS in 2008 with weights gradually shifting towards OIS by the end of 2008.

Further, we use the notation ℓ_t^n for yields corresponding to hypothetical zero-coupon bond prices L_t^n. Such prices can be inferred from quoted Libor rates, $\ell_t^{q,n}$, via $L_t^n = (1 + \ell_t^{q,n} \cdot n \cdot 30/360)^{-1}$ for $n \leq 12$. As a result, although we refer to ℓ_t^n as Libor rates, they are different but close.

The data on forward exchange rates come from Barclays and has maturities $1, 3, 6$ and 12 months. The currency forward data implies, via CIP, interest rate rate differentials $\Delta_c \ell_t^n = \ell_t^n - \widehat{\ell}_t^n$ for the corresponding maturities. By imposing CIP, we are inferring an implicit foreign bank funding rate as opposed to an observable quantity. Such an interpretation is valid in the light of research focusing on various market frictions leading to violations of CIP in terms of actual Libor rates (e.g., Borio, McCauley, McGuire, and Sushko, 2016).

As discussed in Section 2, all foreign government zero-coupon yields are downloaded from their respective central banks (U.S. Federal Reserve, Bank of England, Bundesbank, Bank of Canada and the Bank of Japan). We have maturities of $12, 24, 36, 48, 60, 72, 84, 96, 108, 120$ months for all five countries. Also, we observe the 3 month yield for the U.S. and United Kingdom. Price level data are from the OECD.

We use bilateral data on the U.S. and a foreign country that include depreciation rate, inflation differential, LIBOR and governement interest rates of both countries to estimate the model. The model is cast in a state-space form and is estimated using Bayesian MCMC. See Appendix F.

5 Results

5.1 Initial observations

We report the estimated parameters in Tables 1-4. The first row of each table shows how the expected depreciation rate loads on the different state variables. All of them seem to matter for predictions for the following period, although $\Delta_c \ell_t$ and e_t appear to be particularly significant. We will evaluate the relative importance of the variables for forecasting at different horizons in the subsequent sections. Some of the variables are close to having a unit root under the risk-adjusted probability, but the overall system is stationary (the largest eigenvalue of Φ_x^* is less than one).

The model fit is good. Table 5 displays yield fitting errors. They range between 12 and 57 basis points (on an annualized basis).

The model is also successful in replicating country-specific patterns that were documented in Figure 1. Indeed, Figure 3 illustrates how both actual and risk-adjusted forecasting patterns in the model are capable of capturing the respective pattern in the data. Note that, while the regression models that generate Figure 1 focus on one aspect of evidence at a time (true expectations of depreciation rates, risk-adjusted expectations of the same), our no-arbitrage model captures many aspects of the evidence jointly. Those include true and risk-adjusted expectations of depreciation rates, US and foreign yields, and macro variables.

The covariances of risk-adjusted distribution are relatively precisely estimated with tight highest posterior density intervals, which is typical for no-arbitrage models. Estimates of the covariances are more uncertain under the actual distribution. The amount of uncertainty is consistent with that of regression-based evidence in Figure 1 and reflects the general difficulty of estimating expected returns.

5.2 PPP/cointegration

In general, coefficients α_f and α_f^* appear to be small. Their impact is determined by the product of a specific parameter and the real exchange rate which is much more volatile than the other elements of the state x_t. For convenience, Table 6 summarizes the estimates of α_f's and their risk-adjusted counterparts after re-scaling all the elements in the state vector by their unconditional volatility.

Very few values are large even after rescaling. Parameters α_s and α_π appear to be important across all countries. The risk-adjusted α_π^* is larger than its counterpart under the true probability ($\alpha_s^* = 0$ because of CIP). All other values of α_f^* are smaller than their counterparts. In light of these observations and the requirements outlined in section 2.4, we see that non-monotonicity arises via e_t forecasting Δs_t (non-zero α_s).

These results imply that the variance of the SDF loads on the RER with significant coefficients. Specifically, $\lambda_{e,s} = \alpha_s - \alpha_s^*$ and $\lambda_{e,\pi} = \alpha_\pi - \alpha_\pi^*$ are negative and significant across all countries. Therefore, the maximum risk premium increases when the value of the US consumption basket appreciates.

Further, we quantify the contribution of the RER to the volatility of the SDF by reporting the ratio of the variance of the RER component of the risk premium to the overall variance of the risk premium. The vector of risk premiums λ_t has nine elements. We are focusing on the first two that correspond to the currency and inflation differential premiums, because the respective α's are large and significant. For the depreciation rate, the variance ratio ranges from 6.4% for Canada to 25.9% for the UK. This indicates an economically large contribution of the RER to the maximum currency risk premium. For the inflation differential, the largest variance ratio is 1% (for Japan), so the most important contribution of the RER comes through the currency premium channel.

These observations deepen the exchange rate disconnect (see, e.g., Itskhoki and Mukhin, 2017 for a recent overview). Risk premiums are known to be countercyclical, while there is weak empirical and theoretical connection of the RER to business cycles. Thus, the aforementioned relationship captures quantitatively important movement in risk premiums that is unrelated to the basic macroeconomic aggregates. In section 6 we demonstrate the equilibrium-based modeling challenges in the context of the workhorse asset-pricing models.

Are there other stationary variables besides e_t that could generate this monotonicity? Evidently, not through the same channel as there are no other variables in our model that predict Δs_t in a significant way. But, there are variables that predict $\Delta_c \ell_t$ and are predicted by it. Examples are differences in slopes: $\Delta_c \ell_t^{12,1}$ for the U.K., or $\Delta_c y_t^{120,12}$ for Euro and Japan.

We argue that these variables cannot be solely responsible for the non-monotonic pattern in γ^n. One argument is based on additional multi-horizon evidence motivated by the real exchange rate. Our second argument is based on a VAR model that does not include the real exchange rate, but is otherwise equivalent to the VECM model that we have discussed so far.

5.3 Additional evidence

Results in Dahlquist and Penasse (2016) and our model suggest that e_t is a strong predictor of Δs_t. We extend this result by implementing the UIP-style regressions of section 2.3, but where the IRD is replaced by the RER. Eichenbaum, Johannsen, and Rebelo (2018) explore similar regressions. Figure 4 presents the results.

There is a strong pattern of predictability of nominal depreciation rates via RER across horizons. In contrast, the risk-adjusted regression produces coefficients that are close to zero. This result suggests that the RER is approximately unspanned by forward nominal exchange rates.

Our model can replicate this pattern as the same Figure indicates. Obviously, the pattern under the real-world probability cannot be replicated by a model without the RER. Thus, the evidence reinforces the need to include the RER in our model. The pattern under the risk-adjusted probabilities is, indeed, obtained due to a nearly unspanned RER in forward nominal exchange rates.

To see how that works, recall that the (log) forward exchange rate is equal to the difference between the domestic and foreign yields:

$$fs_t^{n-1} = fy_t^{n-1} - \widehat{fy}_t^{n-1} = n(y_t^n - \widehat{y}_t^n) - (n-1)\,(y_t^{n-1} - \widehat{y}_t^{n-1}).$$

As a result, bond pricing formulas in Appendix D imply that the loadings of n-period forward exchange rates on the factors x_t are equal to $\Phi_x^{*n\top}\delta_{s,x}$. This conclusion holds regardless of the reference curve: Libor-based or government. Because $\delta_{s,x} = \mathrm{e}_1$ in our parametrization, the RER is unspanned in the forward exchange rate curve if the last element of the first row of Φ_x^{*n} is equal to zero for any n.

For instance, this happens if $\alpha_f^* = 0$, similar to Duffee (2011). That's an intriguing possibility because if $\alpha_f^* \approx 0$, the RER is approximately non-stationary under the risk-adjusted probability. Such risk-adjusted values would reflect compensation for market participants who take implicit positions in mean-reverting real exchange rates, but fear that real exchange rates will not revert, or the reversion would take a much longer time than expected. However, in our case α_π^* is economically different from zero.

There is an alternative way to achieve a nearly unspanned RER. When $n = 1$, the last element of the first row of Φ_x^{*n} is equal to α_s^*, which is equal to zero by CIP. In both our simple model of section 2.4 and our full model, this element is equal to α_ℓ^* when $n = 2$. Empirically, α_ℓ^* is close to zero. In the simple model of section 2.4, a value of α_ℓ^* close to zero guarantees that the condition holds approximately for longer maturities n as well. Because $\alpha_\pi^* \neq 0$, we would also need $\phi_{\ell\pi}^* = 0$. This is the case in our simple model by assumption. In the larger model element $\Phi_{x62}^* \equiv \phi_{\ell\pi}^*$ and is estimated to be close to zero.

Does this result imply that the RER is a factor unspanned by the U.S. or foreign bonds? Not necessarily. The conditions above ensure that loadings of domestic and foreign bonds on the RER are the same. But this does not imply that they are equal to zero. For the RER to be unspanned by bonds, we need extra restrictions on the exposure of the spot interest rate to the factors.

These translate into $\Phi_{xk2}^* = 0$ (interaction between kth element of x_t and $\Delta_c\pi_t$) for all k with the exception of $k = 2$ (the diagonal element) and $k = 9$ (the element corresponding to e_t because it is connected to Φ_{x22}^* via cointegrating restrictions). These conditions hold approximately in the estimated model. We confirm that e_t is approximately unspanned by yields by regressing yields on the elements of x_t. These results are available upon request.

If the RER is unspanned by yields does it help in predicting excess bond returns in the spirit of Cochrane and Piazzesi (2005)? We run two types of regressions of the bond excess return on the RER with and without the CP factor. Without the CP factor, the RER does affect bond risk premiums. However after controlling for the CP factor, the predictive ability of RER is eliminated.

5.4 Comparison to a model without cointegration

We compare the VECM model to a model with VAR dynamics that does not include the real exchange rate. Other than dropping the real exchange rate, everything else is the same as in Section 4.3, that is, f_t is unchanged. This model is equivalent to imposing the restriction $\alpha_f = \alpha_f^* = 0$ in the larger VECM, implying that real exchange rates are non-stationary. After imposing the restriction, we re-estimate the model to ensure the best possible fit. Figure 5 plots the UIP regression slopes as a function of horizon for both the VAR and VECM models. The VAR model is clearly incapable of generating a non-monotonic pattern.

5.5 Long-horizon UIP

Chinn and Meredith (2004) propose to test long-horizon UIP using regressions similar to (6). Under UIP, the average depreciation rate between t and $t+n$ should be explained by the difference in n maturity yields across countries.

$$n^{-1}\sum_{j=1}^{n}\Delta s_{t+j} = \tilde{\gamma}_0^n + \tilde{\gamma}^n \Delta_c y_t^n + \tilde{u}_{t+n}, \qquad n = 1, 2, \ldots, 120. \tag{35}$$

UIP predicts that $\tilde{\gamma}^n = 1$ for any horizon n. Chinn and Meredith (2004) find that UIP holds approximately at longer horizons of $n = 60$ and 120 months. We replicate their finding in Figure 6. In fact, because we investigate a broader spectrum of horizons, we document a non-monotonic pattern akin to the one in regression (6): regression slopes between the Chinn-Meredith horizons of 5 and 10 years are larger than one, albeit not statistically signifcant.

The two regressions must be related because the left-hand side in (35) is just an aggregation of that in (6). There is also a similar no-arbitrage relationship explaining why there should be a bias (deviation from 1):

$$\Delta_c y_t^n = n^{-1}\log E_t^*\left[\exp\left(\sum_{j=1}^{n}\Delta s_{t+j}\right)\right],$$

which implies a slope of 1 under the risk-adjusted probability if the state variables are homoscedastic.

The difference between the regressions in (35) and (6) is that the regressor's maturity is shifting with horizon, which has two implications. First, analysis of regressions in (35) requires a model of risk premiums such as ours. Second, the risk-adjusted slope is equal to 1 for any n, and not just $n = 1$. That's why we do not report a counterpart to regression (7). The issue is then whether a model can replicate the pattern in risk premiums that is responsible for the pattern observed in the data. As we've seen, risk premiums take a different form depending on whether PPP holds or not.

Figure 6 shows patterns implied by both the VECM and VAR models. The VECM is closer to the data. Most important, while the VAR settles at $\tilde{\gamma}^{120}$ of about zero, the VECM crosses into the positive territory at horizon $n = 24$.

5.6 The currency premiums

Figure 1 implicitly tells us about the nominal FX risk premium in (3). Because $\Delta s_t = \mathrm{e}_1^\top f_t$, we can compute these risk premiums using the same techniques as the ones used for bond prices. In particular,

$$rps_t^1 = \mathrm{e}_1^\top \left[\mu_f - \mu_f^* + (\Phi_f - \Phi_f^*) f_t + (\alpha_f - \alpha_f^*) e_t \right].$$

We use the notation

$$rps_t^n = \kappa_{n,0} + \kappa_{n,x}^\top x_t$$

to simply refer to the factor loadings $\kappa_{n,x}$ for longer horizons.

As we noted earlier, the terms on the right hand side are equal, up to convexity, to $E_t \left[\Delta s_{t+n} \right]$ and $E_t^* \left[\Delta s_{t+n} \right]$, respectively. Figure 1 shows coefficients corresponding to a projection of these risk premiums onto $\Delta_c \ell_t$. Thus, the difference between the two lines multiplied by $\Delta_c \ell_t$ produces a projection of rps_t^n on the IRD.

We can compare this projection to the full risk premium implied by the model. Before computing this risk premium, we asses the statistical significance of the factor loadings $\kappa_{n,x}$. Following our analysis of the UIP regression coefficients, we display the average of country-specific $\kappa_{n,x}$ in Figure 7. To save space, we show only the loadings that are significant at least at some horizons. We see that, in addition to the IRD and the RER, one should be incorporating the differential in term spreads $\Delta_c y_t^{120,12}$ and the difference in term spreads of the Libor curve $\Delta_c \ell_t^{12,1}$. Observe that, while the IRD and $\Delta_c \ell_t^{12,1}$ are significant at all horizons, the RER becomes insignificant after 60 months, while $\Delta_c y_t^{120,12}$ becomes significant after 60 months.

Consistent with our discussion in section 3.1 the pattern of loadings on IRD is non-monotonic despite the inclusion of the RER and other variables. That is precisely because α_s and α_π are significant and, as a result, generate the non-monotonicity. Thus, the mere fact of PPP is responsible for the documented patterns associated with the IRD. The predictability of depreciation rates with the RER is a bonus.

On the basis of these results, we construct risk premiums rps_t^n using only the four factors with significant loadings, and multiplying the country-specific factors by the cross-country averages of $\kappa_{n,x}$ from Figure 7. Figure 8 compares rps_t^n constructed this way to its projection on the IRD. We find that the projected version is less variable. While, mathematically, this result is to be expected, the numerical difference is quite large. UIP regressions appear to leave a lot out in terms of risk premium measurement.

Besides the scale, the two versions can be quite different at times. The standard intuition is that the risk premium moves in the direction opposite to the IRD. Here we observe that quite often the projection and the full risk premium move in opposite directions implying that the effect of the IRD is overwhelmed by the other three variables.

This evidence adds a new dimension to the UIP regressions. Not only does UIP not hold at different horizons, but deviations from UIP are driven not by IRDs alone. Future research should consider the role of term spreads in determination of currency risk premiums.

6 Models of exchange rates with exogenous consumption under complete markets

In this section, we explain why workhorse consumption-based asset pricing models have a difficult time simultaneously replicating the qualitative features of the data including (i) a stationary RER; (ii) more than one predictor of currency excess returns; (iii) and a RER that enters the conditional variance of the SDF. We provide all the derivations in Appendix G. Here we outline the main logical steps leading to this conclusion.

First, we limit our study to open-economy models with an exogenous stream of consumption in each country. Formally, in this setting there is no trade in real assets. Such a framework could be viewed as partial equilibrium with the relationship between consumption and output unmodeled. The hope is that one could support the assumed consumption processes by an explicit model of trade. For the RER to vary, such a model should feature either differences in preferences across countries or trade frictions in goods markets. As we will see, the former requirement takes a back seat in or setting because of the stationarity of the RER. Thus, one would need a model featuring the latter requirement to support our assumptions. In order to be qualitatively consistent with trade frictions, we will assume different consumption dynamics in different countries. Further, we assume complete financial markets, which allows computation of the real depreciation rate as the ratio of the real SDFs in two countries.

There is virtually an infinite number of specifications that one could entertain and, that is why we limit the scope of our analysis by using existing models that match the characteristics discussed above as an organizing principle. Specifically, we specify exogenous consumption that resembles processes in Bansal and Shaliastovich (2013), Colacito and Croce (2011), and Verdelhan (2010). This research explores different questions, so our intent is not to critique these specific papers but to understand the framework's limitations in the context of our empirical findings.

Second, stationarity of the RER depends on the long-run properties of the real SDF (Alvarez and Jermann, 2005, Hansen and Scheinkman, 2009). Specifically, as shown by Backus, Boyarchenko, and Chernov (2018) and Lustig, Stathopoulos, and Verdelhan (2013), the martingale, or permanent, components of the domestic and foreign real SDFs must coincide for the RER to be stationary. Equivalently, the martingale component of the RER has to be constant.

Third, as shown by Backus, Chernov, and Zin (2014), Bansal and Lehmann (1997), and Hansen (2012), the long-horizon properties of models with recursive or habit-based preferences are similar to those of the model with additive power utility. Thus, we develop restrictions needed for the stationarity of the RER in the context of a model with power utility. Then, we apply these restrictions to other preference specifications.

Fourth, a necessary and sufficient condition for stationary of the RER in this context is that the (log) consumption processes in the two countries are cointegrated. To show this, we use MU_t to denote the marginal utility of the domestic representative agent, so that the real SDF is $\mathcal{M}_{t,t+1} = MU_{t+1}/MU_t$ (in this section we have repurposed the notation $\mathcal{M}$ from the nominal to the real SDF). Thus, in the power utility case, assuming the same time preference β for both countries,

$$\frac{\widehat{MU}_t}{MU_t} = e^{(\widehat{\alpha}-1)\widehat{c}_t-(\alpha-1)c_t}, \tag{36}$$

where c_t and $\widehat{c}_t$ denotes the log of US and foreign aggregate consumption, respectively. The respective coefficients of relative risk aversion are $1-\alpha$, and $1-\widehat{\alpha}$. If $(\widehat{\alpha}-1)\widehat{c}_t-(\alpha-1)c_t$ is (non-) stationary, then $\widehat{MU}_t/MU_t$ is (non-) stationary, and its growth rate, $\widehat{\mathcal{M}}_{t,t+1}/\mathcal{M}_{t,t+1}$ will (not) have a constant martingale / permanent component. That would imply a (non-) stationary RER.

Fifth, we assume a specific form of cointegration for consumption across countries. Equation (36) suggests that $(\widehat{\alpha}-1, 1-\alpha)^\top$ is the cointegrating vector. However, it appears to be far-fetched to link cointegration to preferences, at least without an explicit model of trade. Thus, we make a simplifying assumption that the cointegrating vector is $(1,-1)^\top$. Thus the variable $\Delta_c c_t \equiv c_t - \widehat{c}_t$ is stationary. This assumption implies that risk aversion should be the same, $\alpha = \widehat{\alpha}$.

Sixth, we specify the dynamics of consumption growth across countries to follow a (restricted) VECM process:

$$\begin{aligned}
\Delta c_{t+1} &= \mu + \theta \Delta_c c_t + \sigma_t \varepsilon_{t+1}, \\
\Delta \widehat{c}_{t+1} &= \widehat{\mu} + \widehat{\theta} \Delta_c c_t + \widehat{\sigma}_t \widehat{\varepsilon}_{t+1}, \\
\Delta_c c_{t+1} &= \Delta_c c_t + \Delta c_{t+1} - \Delta \widehat{c}_{t+1} \\
&= \mu - \widehat{\mu} + (1+\theta-\widehat{\theta})\Delta_c c_t + \sigma_t \varepsilon_{t+1} - \widehat{\sigma}_t \widehat{\varepsilon}_{t+1}.
\end{aligned}$$

This specification is new to the literature as all existing models in this class feature a non-stationary RER. Nevertheless, it resembles the aforementioned studies. The time-varying volatility of consumption growth (σ_t or $\widehat{\sigma}_t$) is associated with macroeconomic activity at the business-cycle frequency (e.g., Segal, Shaliastovich, and Yaron, 2015).

In a model with power utility, the RER is

$$\mathcal{E}_t = \mathcal{E}_0 \widehat{\mathcal{M}}_{0,t}/\mathcal{M}_{0,t} = \mathcal{E}_0 e^{(\alpha-1)\Delta_c c_0} \cdot e^{(1-\alpha)\Delta_c c_t}. \tag{37}$$

The domestic and foreign real interest rates are, respectively,

$$\begin{aligned} r_t &= -\log\beta + (1-\alpha)E_t\left[\Delta c_{t+1}\right] \\ \hat{r}_t &= -\log\beta + (1-\alpha)E_t\left[\Delta \widehat{c}_{t+1}\right]. \end{aligned}$$

Thus, the interest rate differential

$$\Delta_c r_t = (1-\alpha)E_t\left[\Delta c_{t+1} - \Delta \hat{c}_{t+1}\right] = (1-\alpha)[\mu - \widehat{\mu} + (\theta - \widehat{\theta})\Delta_c c_t] \tag{38}$$

summarizes the same information as the RER and one cannot obtain an extra forecasting variable for the excess currency return.

To demonstrate this point, we need to switch to nominal variables. Given illustrative purposes of the exercise, we follow Bansal and Shaliastovich (2013) and assume an exogenous process for the inflation differential. We use the simple model of section 3:

$$\Delta_c \pi_{t+1} = \mu_\pi + \phi_\pi \Delta_c \pi_t + \alpha_\pi e_t + \sigma_\pi \varepsilon_{\pi t+1}.$$

Combining the dynamics of inflation differentials with changes in the real exchange rate implied by (37) and (38), the resulting nominal depreciation rate is

$$\begin{aligned} \Delta s_{t+1} &= \Delta e_{t+1} + \Delta_c \pi_{t+1} \\ &= (1-\alpha)(\mu - \widehat{\mu}) + \mu_\pi + e_0 + (\alpha - 1)\Delta_c c_0 \\ &+ (1-\alpha)(\theta - \widehat{\theta})\Delta_c c_t + \alpha_\pi (1-\alpha)\Delta_c c_t + \phi_\pi \Delta_c \pi_t \\ &+ \sigma_t \varepsilon_{t+1} - \widehat{\sigma}_t \widehat{\varepsilon}_{t+1} + \sigma_\pi \varepsilon_{\pi t+1}. \end{aligned}$$

Thus, the log expected excess return is

$$E_t\left[\Delta s_{t+1} - (\Delta_c r_t + \Delta_c \pi_{t+1})\right] = e_0 + (\alpha - 1)\Delta_c c_0 - \alpha_\pi (1-\alpha)\Delta_c c_t$$

and there is only one forecasting variable, which could be extracted either from the IRD or the RER. If $\alpha_\pi = 0$, as would be the case in a more traditional analysis, the expected excess return would be constant.

The conditional variance of the log SDF

$$var_t(\log \mathcal{M}_{t,t+1}) = (\alpha - 1)^2 \sigma_t^2$$

is not related to the RER. Consequently, it is impossible for these models to generate a time-varying risk premium that is a function of the RER. We show in Appendix G that extensions of the power utility model to recursive or habit-based preferences continue having challenges with matching our evidence on predictability and variation in the SDF.

Engel (2016) rejects models with recursive preferences on the basis of the multi-horizon UIP pattern alone. However, the models he evaluates feature non-stationary real exchange rates, which, of course, would not be able to capture the UIP evidence. Therefore, once one accounts for stationarity of the RER, the UIP-based evidence could be insufficient for discriminating between the models. That observation highlights the importance of establishing SDF dynamics for the purposes of model diagnostics.

7 Conclusion

Exposures of expected future depreciation rates to the current interest rate differential violate the UIP hypothesis across horizons in a distinctive pattern that is a non-monotonic. Conversely, forward, risk-adjusted expected depreciation rates are monotonic. We offer a potential explanation for why these patterns occur. The non-monotonicity is driven by the presence of the real exchange rate (PPP deviations) as a predictive variable in addition to the interest rate differential appearing in the UIP regressions. The real exchange does not affect risk-adjusted expectations because of the risk premium for departures from PPP. The existence of this risk premium implies that conditional variance of the stochastic discount factor, a.k.a. maximal risk premium depends on the real exchange rate.

To illustrate this mechanism, we built a no-arbitrage term structure model with VECM dynamics that includes the real exchange rate as a state variable. Including state variables that are cointegrated into the dynamics of the model is new to the literature on no arbitrage term structure models. Estimates from the model provide evidence that supports our explanation. Finally, we show that existing equilibrium models with exogenous consumption have difficulty generating a stochastic discount factor with the documented dependence on the real exchange rate.

References

Alvarez, Fernando, and Urban J. Jermann, 2005, Using asset prices to measure the persistence of the marginal utility of wealth., *Econometrica* 73, 1977–2016.

Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745–787.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *The Journal of Finance* 68, 929–985.

Bacchetta, Philippe, and Eric van Wincoop, 2010, Infrequent portfolio decisions: A solution to the forward discount puzzle., *American Economic Review* 3, 870–904.

Backus, David K., Nina Boyarchenko, and Mikhail Chernov, 2018, Term structures of asset prices and returns, *Journal of Financial Economics* 129, 1–23.

Backus, David K., Mikhail Chernov, and Stanley Zin, 2014, Sources of entropy in representative agent models, *The Journal of Finance* 69, 51–99.

Balduzzi, Pierluigi, and I-Hsuan Ethan Chiang, 2017, Real exchange rates and currency risk premia, Boston College, Working paper.

Bansal, Ravi, and Bruce Lehmann, 1997, Growth-optimal portfolio restrictions on asset pricing models., *Macroeconomic Dynamicsß* 1, 333–354.

Bansal, Ravi, and Ivan Shaliastovich, 2013, A long-run risks explanation of predictability puzzles in bond and currency markets, *The Review of Financial Studies* 26, 1–33.

Barroso, Pedro, Sara Filipe, and Paulo F. Maio, 2016, What drives exchange rates? reassessing currency return predictability, Working Paper, Luxembourg School of Finance.

Bilson, John F. O., 1981, The speculative efficiency hypothesis., *Journal of Business* 54, 435–451.

Binsbergen, Jules van, and Ralph Koijen, 2017, The term structure of returns: facts and theory, *Journal of Financial Economics* 124, 1–21.

Borio, Claudio, Robert Neil McCauley, Patrick McGuire, and Vladyslav Sushko, 2016, Covered interest parity lost: understanding the cross-currency basis, *BIS Quarterly Review.*

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw, 2016, New evidence on the forward premium puzzle, *Journal of Financial and Quantitative Analysis* 51, 875–897.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: a consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Chinn, Menzie, and Guy Meredith, 2004, Monetary policy and long-horizon uncovered interest parity, *IMF staff papers* 51, 409–430.

Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *The Review of Financial Studies* 21, 1533–1575.

——— , and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Colacito, Riccardo, and Mariano M. Croce, 2011, Risks for the long run and the real exchange rate., *Journal of Political Economy* 119, 153–181.

Dahlquist, Magnus, and Julien Penasse, 2016, The missing risk premium in exchange rates, Stockholm School of Economics, Working Paper.

Du, Wenxin, Joanne Im, and Jesse Schreger, 2018, The U.S. Treasury premium, *Journal of International Economics* 112, 167–181.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan, 2016, Deviations from covered interest parity, Unpublished manuscript, MIT Sloan.

Duffee, Gregory R., 2011, Information in (and not in) the term structure, *The Review of Financial Studies* 24, 2895–2934.

Duffie, Darrell, and Kenneth J Singleton, 1999, Modeling term structures of defaultable bonds, *The Review of Financial Studies* 12, 687–720.

Eichenbaum, Martin, Benjamin K. Johannsen, and Sergio Rebelo, 2018, Monetary policy and the predictability of nominal exchange rates, working paper.

Engel, Charles, 2016, Exchange rates, interest rates, and the risk premium, *American Economic Review* 106, 436–474.

Engle, Robert F, and Clive W. J. Granger, 1987, Co-integration and error correction: representation, estimation and testing, *Econometrica* 55, 251–276.

Fama, Eugene F., 1984, Forward and spot exchange rates., *Journal of Monetary Economics* 14, 319–338.

——— , and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Gurkaynak, Refet, Brian Sack, and Jonathan Wright, 2007, The U.S. Treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291–2304.

Hamilton, James D, 1994, *Time Series Analysis* (Princeton University Press: Princeton, NJ).

Hansen, Lars Peter, 2012, Dynamic valuation decomposition within stochastic economies., *Econometrica* 80, 911–967.

——— , and José A. Scheinkman, 2009, Long-term risk: An operator approach., *Econometrica* 77, 177–234.

Itskhoki, Oleg, and Dmitry Mukhin, 2017, Exchange rate disconnect in general eqilibrium, working paper.

Jiang, Zhengyang, Arvind Krishanmurthy, and Hanno Lustig, 2018, Foreign safe asset demand and the dollar exchange rate, working paper.

Jorda, Oscar, and Alan M. Taylor, 2012, The carry trade and fundamentals: Nothing to fear but FEER itself, *Journal of International Economics* 88, 74–90.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, 2011, A new perspective on Gaussian affine term structure models, *The Review of Financial Studies* 27, 926–970.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan, 2013, The term structure of currency carry trade risk premia, working paper.

Meese, Richard A., and Kenneth Rogoff, 1983, Empirical exchange rate models of the seventies: Do they fit out of sample?, *Journal of International Economics* 14, 3 – 24.

Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf, 2017, Currency value, *The Review of Financial Studies* 30, 416–441.

Segal, Gill, Ivan Shaliastovich, and Amir Yaron, 2015, Good and bad uncertainty: Macroeconomic and financial market implications, *Journal of Financial Economics* 117, 369–397.

Tryon, Ralph, 1979, Testing for rational expectations in foreign exchange markets, Federal Reserve Board international finance discussion paper no. 139.

Valchev, Rosen, 2016, Bond convenience yields and exchange rate dynamics, Department of Economics, Boston College, Working Paper.

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

Appendix A An extra variable affecting joint dynamics of IRD and depreciation rate

Appendix A.1 VAR representaton

A natural starting point for thinking about joint dynamics of the depreciation rate Δs_t and the IRD $\Delta_c \ell_t$ is a simple vector autoregression. We would like to highlight properties of another generic variable v_t that affects these dynamics. Specifically, our focus is on what properties does a simple VAR model that includes v_t need to have in order to generate the patterns in γ^n and γ^{*n} that were documented in Section 2. Simultaneously, under the risk-neutral distribution, the same coefficients must be monotonic and of opposite sign. To keep ideas tractable, we focus on the case where v_t is univariate.

Stack the state variables into a vector x_t: $x_t = (\Delta s_t \ \Delta_c \ell_t \ v_t)^\top$. To simplify the setup, we ignore means (assume all variables have mean zero) and model the state vector as a first order process

$$x_t = \Phi_x x_{t-1} + \Sigma_x \varepsilon_t$$

with

$$\Phi_x = \begin{pmatrix} 0 & \phi_{s\ell} & \phi_{sv} \\ 0 & \phi_\ell & \phi_{\ell v} \\ 0 & \phi_{v\ell} & \phi_v \end{pmatrix}, \quad \Sigma_x = \begin{pmatrix} \sigma_s & 0 & 0 \\ 0 & \sigma_\pi & 0 \\ 0 & 0 & \sigma_v \end{pmatrix}.$$

Our discussion centers on the autocovariance matrix Φ_x which determines the covariances between variables at alternative horizons. In our simple illustration, we set the first column of Φ_x to zero by assumption although this value is empirically realistic. Depreciation rates are not highly autocorrelated and do not forecast the IRD. The coefficient $\phi_{s\ell}$ reflects the UIP regression. If UIP were to hold, we should anticipate coefficients in the first row of Φ_x to be $\phi_{s\ell} = 1$ and $\phi_{sv} = 0$.

The values of γ^n reported in Section 2 are directly related to the forecast function of the VAR. The forecast $E_t[\Delta s_{t+n}]$ is controlled by exponents of the matrix Φ_x

$$E_t[\Delta s_{t+n}] = \mathrm{e}_1^\top \Phi_x^n x_t, \quad \mathrm{e}_1^\top = (1, 0, 0).$$

In general, it is difficult to obtain tractable closed-form expressions for long horizons n. We can do so for $n = 1, 2, 3$ in the case of our simple model:

$$\begin{aligned} E_t[\Delta s_{t+1}] &= \phi_{s\ell} \Delta_c \ell_t + \phi_{sv} v_t, & \text{(A.1)} \\ E_t[\Delta s_{t+2}] &= (\phi_{s\ell}\phi_\ell + \phi_{v\ell}\phi_{sv}) \Delta_c \ell_t + (\phi_{s\ell}\phi_{\ell v} + \phi_{sv}\phi_v) v_t. & \text{(A.2)} \\ E_t[\Delta s_{t+3}] &= \left(\phi_{s\ell}\left(\phi_\ell^2 + \phi_{\ell v}\phi_{v\ell}\right) + \phi_{sv}\left(\phi_{v\ell}\phi_\ell + \phi_{v\ell}\phi_v\right)\right) \Delta_c \ell_t & \\ &\quad + \left(\left(\phi_{s\ell}\phi_{\ell v}\left(\phi_\ell + \phi_v\right) + \phi_{sv}\left(\phi_v^2 + \phi_{v\ell}\phi_{\ell,s}\right)\right) v_t. & \text{(A.3)} \end{aligned}$$

In these expressions, the loadings on the IRD $\Delta_c \ell_t$ have the largest impact on the coefficients γ^n. At horizon $h = 1$, the covariance γ^1 is a function of only the UIP coefficient $\phi_{s\ell}$. Because $\phi_{s\ell}$ is typically estimated as large and negative, the covariance γ^1 is negative, which is consistent with the patterns documented in Section 2.

As the horizon increases, the loadings on $\Delta_c \ell_t$ can be written as the sum of two terms that are controlled by the forecasting parameters $\phi_{s\ell}$ and ϕ_{sv}. We highlight these here for $n = 3$ as

$$\begin{aligned} \text{term 1} &= \phi_{s\ell}\left(\phi_\ell^2 + \phi_{\ell v}\phi_{v\ell}\right) \\ \text{term 2} &= \phi_{sv}\left(\phi_{v\ell}\phi_\ell + \phi_{v\ell}\phi_v\right) \end{aligned}$$

The first term contains $\phi_{s\ell}$ and it multiplies powers of the IRD autocorrelation coefficient ϕ_ℓ^2, which becomes ϕ_ℓ^{n-1} at longer horizons. This term induces a slow monotonic decay in the covariances as the horizon increases and it is the dominant component of γ^n, especially at short horizons. If the forecasting variable v_t were not present in the model ($\phi_{sv} = 0, \phi_{\ell v} = 0$), then the cross auto-covariance between the depreciate rate and the IRD would simply decay monotonically because it is influenced only by the product $\phi_{s\ell}\phi_\ell^{n-1}$ as the horizon increases. We conclude that a first-order VAR with only the depreciation rate and IRD would not generate the non-monotonic pattern we observe in practice.

Appendix A.2 Non-monotonic pattern in γ^n

Next, we will illustrate how this model can generate non-monotonic patterns through two possible channels. Although it is possible that both could be present simultaneously, we illustrate them one at a time. In the first channel, the variable v_t may forecast the depreciation rate, $\phi_{sv} \neq 0$, while having no impact on the IRD itself, $\phi_{\ell v} = 0$. The second possible channel occurs if the variable v_t forecasts the IRD, $\phi_{\ell v} \neq 0$, while it does not forecast the depreciation rate $\phi_{sv} = 0$.

If the first channel is at play, term 2 in the analytical expression for $n = 3$ above starts small at short horizons but begins to dominate term 1 at intermediate horizons before the system as a whole converges back to equilibrium.

In the second case term 2 has no influence. Instead, the loading on the IRD is a function of term 1 only. One component of the loading contains a power, ϕ_ℓ^2, which induces monotonocity. Another component, $\phi_{\ell v}\phi_{v\ell}$, can induce non-monotonicity. As the horizon increases, this second component must be large enough to dominate the monotonic component.

Finally, the IRD must forecast the variable v_t, $\phi_{v\ell} \neq 0$, for either channel to work. If it does not, then the cross-autocovariances are monotonic no matter what the values of ϕ_{sv} and $\phi_{\ell v}$ are. This is clear from the analytical expressions for horizons $h = 2, 3$ shown above, and we illustrate this numerically below.

Appendix A.3 Monotonic pattern in γ^{*n}

This discussion has an immediate implication for the risk-adjusted dynamics of the state x_t should follow a VAR in order to replicate the monotonic pattern of γ^{*n} in Figure 1. Under risk-adjusted probability, the persistence matrix Φ_x^* could be different from Φ_x, including the zero elements becoming non-zero. For the purposes of this discussion we simplify and assume the following form:

$$\Phi_x^* \quad = \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & \phi_\ell^* & \phi_{\ell,v}^* \\ 0 & \phi_{v,\ell}^* & \phi_v^* \end{pmatrix}.$$

The first row is dictated by the fact that UIP must hold in the risk-adjusted world.

The UIP-imposed restrictions that $\phi_{s\ell}^* = 1$ and $\phi_{sv}^* = 0$ already suggests that the risk-adjusted pattern could have different properties. First, the first-order cross-autocovariance $\mathrm{e}_1^\top \Phi_x^* \mathrm{e}_2$ must be equal to one, consistent with the evidence. Second, the restriction $\phi_{sv}^* = 0$ rules out the possibility of inducing non-monotonic patterns in the auto-covariances through the first channel. If this is the channel that induces the real world covariances to be non-monotonic, it has implications for currency risk premia.

Appendix B Relationship to a standard VECM(1,1)

In this appendix, we explain how our error correction model of cointegration differs from a traditional VECM. Ultimately, the models are equivalent but parameterized in different ways. First, a VECM is typically written in terms of a mixture of $I(1)$ and $I(0)$ variables, whereas we write it entirely in terms of $I(0)$ variables. Secondly, we define the matrix of cointegrating vectors to be only those vectors whose linear combinations include non-stationary series. To illustrate these differences, we provide a simple example.

Consider the VECM of Engel (2016). The state vector z_t includes observables s_t and $\Delta_c p_t$ that are $I(1)$ while the interest rate differential $\Delta_c \ell_t$ is $I(0)$.

$$z_t = \begin{pmatrix} s_t \\ \Delta_c p_t \\ \Delta_c \ell_t \end{pmatrix}$$

The vector z_t is a mixture of $I(0)$ and $I(1)$ variables. Taking first differences of z_t, the dynamics of a traditional VECM(1,1) are

$$\Delta z_t = \mu_z + \Gamma_z \Delta z_{t-1} + \Omega_z z_{t-1} + \Sigma_z \varepsilon_t$$

where $\Omega_z = \Psi_z \beta_z^\top$ and β_z is the matrix of cointegrating vectors. Recalling that π_t denotes inflation, we can express this more explicitly in matrices

$$\begin{aligned} \begin{pmatrix} \Delta s_t \\ \Delta_c \pi_t \\ \Delta\Delta_c \ell_t \end{pmatrix} &= \begin{pmatrix} \mu_s \\ \mu_p \\ \mu_\ell \end{pmatrix} + \begin{pmatrix} \gamma_s & \gamma_{s,\pi} & \gamma_{s,\ell} \\ \gamma_{\pi,s} & \gamma_\pi & \gamma_{\pi,\ell} \\ \gamma_{\ell,s} & \gamma_{\ell,\pi} & \gamma_\ell \end{pmatrix} \begin{pmatrix} \Delta s_{t-1} \\ \Delta_c \pi_{t-1} \\ \Delta\Delta_c \ell_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} \psi_{s,e} & \psi_{s,\ell} \\ \psi_{\pi,e} & \psi_{\pi,\ell} \\ \psi_{\ell,e} & \psi_\ell \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_{t-1} \\ \Delta_c p_{t-1} \\ \Delta_c \ell_{t-1} \end{pmatrix} + \Sigma_z \varepsilon_t \end{aligned}$$

It is traditional to include all stationary relationships in the β_z matrix

$$\beta_z^\top = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This means that β_z includes the trivial relationships that are a priori known to be $I(0)$, e.g. see the second row is not a function of any $I(1)$ variables.

Next, we re-write this model using the error correction representation in our paper. The log-likelihoods of these two models are equivalent. First, we express the state vector f_t in terms of only $I(0)$ variables.

$$f_t = \begin{pmatrix} \Delta s_t \\ \Delta_c \pi_t \\ \Delta_c \ell_t \end{pmatrix}$$

Secondly, our model defines the matrix of cointegrating vectors β_f as only a function of the linear combinations that include non-stationary variables.

$$\beta_f^\top = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}$$

We do not include in β_f the "redundant" stationary relationships, i.e. we drop the second row of $\beta_z^\top$ above.

In our notation, the traditional VECM(1,1) above has dynamics.

$$\begin{aligned} \begin{pmatrix} \Delta s_t \\ \Delta_c \pi_t \\ \Delta_c \ell_t \end{pmatrix} &= \begin{pmatrix} \mu_s \\ \mu_p \\ \mu_\ell \end{pmatrix} + \begin{pmatrix} \gamma_s & \gamma_{s,\pi} & \gamma_{s,\ell} + \psi_{s,\ell} \\ \gamma_{\pi,s} & \gamma_\pi & \gamma_{\pi,\ell} + \psi_{\pi,\ell} \\ \gamma_{\ell,s} & \gamma_{\ell,\pi} & 1 + \gamma_\ell + \psi_\ell \end{pmatrix} \begin{pmatrix} \Delta s_{t-1} \\ \Delta_c \pi_{t-1} \\ \Delta_c \ell_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0 & 0 & -\gamma_{s,\ell} \\ 0 & 0 & -\gamma_{\pi,\ell} \\ 0 & 0 & -\gamma_\ell \end{pmatrix} \begin{pmatrix} \Delta s_{t-2} \\ \Delta_c \pi_{t-2} \\ \Delta_c \ell_{t-2} \end{pmatrix} + \begin{pmatrix} \psi_{s,e} \\ \psi_{\pi,e} \\ \psi_{\ell,e} \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} s_{t-1} \\ \Delta_c p_{t-1} \\ \Delta_c \ell_{t-1} \end{pmatrix} + \Sigma_z \varepsilon_t \end{aligned}$$

This shows that a traditional VECM(1,1) has second order dynamics for those variables that were originally stationary, e.g. the interest rate differential $\Delta_c \ell_t$.

In our paper, we set the second-order lag term to zero $\gamma_{s,\ell} = \gamma_{\pi,\ell} = \gamma_\ell = 0$. We could include this in our error correction representation by adding a second lag.

$$f_t = \mu_f + \Phi_{f,1} f_{t-1} + \Phi_{f,2} f_{t-2} + \alpha_f \beta_f^\top f_{t-1}^L + \Sigma_f \varepsilon_t \tag{B.1}$$

We choose not to do this for the benchmark model of our paper.

Appendix C Change of probability

Appendix C.1 Notation

We introduce additional notation that we use throughout the appendix. We define the following set of matrices

$$\begin{aligned}
\mathcal{C} &= \begin{pmatrix} 0 \\ \beta_{f,0} \end{pmatrix} \\
\mathcal{I} &= \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{I} \end{pmatrix} \\
\mathcal{B}_f &= \begin{pmatrix} \mathrm{I} \\ \beta_f^\top \end{pmatrix} \\
\mathcal{A}_f &= \begin{pmatrix} \Phi_f & \alpha_f \end{pmatrix} \\
\Pi_f &= \alpha_f \beta_f^\top \\
S_x &= \Sigma_x \Sigma_x^\top \\
S_f &= \Sigma_f \Sigma_f^\top
\end{aligned}$$

When we state that x_t can be written as a cointegrated system, we mean that the parameters of the vector autoregression

$$x_t = \mu_x + \Phi_x x_{t-1} + \Sigma_x \varepsilon_t$$

can be decomposed as

$$\begin{aligned}
\mu_x &= \mathcal{C} + \mathcal{B}_f \mu_f \\
\Phi_x &= \mathcal{I} + \mathcal{B}_f \mathcal{A}_f \\
\Sigma_x &= \mathcal{B}_f \Sigma_f
\end{aligned}$$

A similar decomposition also holds under the risk-adjusted probability when $(\mu_f, \Phi_f, \alpha_f, \beta_f)$ are replaced by $\left(\mu_f^*, \Phi_f^*, \alpha_f^*, \beta_f^*\right)$.

Appendix C.2 Generalized inverse of $\Sigma_x \Sigma_x^\top$

The matrix $S_x = \Sigma_x \Sigma_x^\top$ is singular. The generalized inverse S_x^+ of S_x is

$$\begin{aligned}
\Sigma_x \Sigma_x^\top S_x^+ \Sigma_x \Sigma_x^\top &= \Sigma_x \Sigma_x^\top \\
\mathcal{B}_f \Sigma_f \left(\mathcal{B}_f \Sigma_f\right)^\top S_x^+ \mathcal{B}_f \Sigma_f \left(\mathcal{B}_f \Sigma_f\right)^\top &= \mathcal{B}_f \Sigma_f \left(\mathcal{B}_f \Sigma_f\right)^\top \\
\Sigma_f^\top \mathcal{B}_f^\top S_x^+ \mathcal{B}_f \Sigma_f &= I_{d_f} \\
\mathcal{B}_f^\top S_x^+ \mathcal{B}_f &= \left(\Sigma_f \Sigma_f^\top\right)^{-1}
\end{aligned}$$

The solution to this equation is

$$S_x^+ \quad = \quad \mathcal{B}_f\left(\mathcal{B}_f^\top\mathcal{B}_f\right)^{-1}\left(\Sigma_f\Sigma_f^\top\right)^{-1}\left(\mathcal{B}_f^\top\mathcal{B}_f\right)^{-1}\mathcal{B}_f^\top$$

We use this below.

Appendix C.3 Prices of risk

Let $\mu_{x,t}$ and S_x denote the conditional mean and covariance matrix of x_t. We define a restriction of Lebesgue measure to the dimension of rank(S_x). The vector x_t has a density w.r.t. to this measure given by

$$p\left(x_{t+1}|x_t;\theta\right) \quad = \quad \det^*\left(2\pi S_x\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(x_{t+1}-\mu_{x,t}\right)^\top S_x^+\left(x_{t+1}-\mu_{x,t}\right)\right)$$

where S_x^+ denotes the generalized inverse and $\det^*$ is the pseudo-determinant.

The stochastic discount factor (SDF) is

$$\mathcal{M}_{t,t+1} \quad = \quad \exp\left(-\ell_t\right)\frac{p\left(x_{t+1}|x_t;\theta^*\right)}{p\left(x_{t+1}|x_t;\theta\right)}$$

Before deriving the SDF, we first write the quadratic form. Using the notation above, the scaled shock can be written as

$$\begin{aligned}
\Sigma_x\varepsilon_{t+1}^* \quad &= \quad \left(x_{t+1}-\mu_{x,t}^*\right) \;=\; \left(\left[\mathcal{B}_f^* f_{t+1}+\mathcal{I}x_t+\mathcal{C}^*\right]-\left[\mathcal{C}^*+\mathcal{B}_f^*\mu_f^*\right]-\left[\mathcal{I}+\mathcal{B}_f^*\mathcal{A}_f^*\right]x_t\right)\\
&= \quad \left(\mathcal{B}_f^* f_{t+1}-\mathcal{B}_f^*\mu_f^*-\mathcal{B}_f^*\mathcal{A}_f^* x_t\right) \;=\; \mathcal{B}_f^*\left(f_{t+1}-\mu_f^*-\mathcal{A}_f^* x_t\right)\\
&= \quad \mathcal{B}_f^*\left(f_{t+1}-\mu_f^*-\Phi_f^* f_t-\Pi_f^* f_t^L\right)
\end{aligned}$$

Plugging this into the quadratic form, we find

$$\begin{aligned}
\left(x_{t+1}-\mu_{x,t}^*\right)^\top S_x^{*,+}\left(x_{t+1}-\mu_{x,t}^*\right) \quad &= \quad \left(\mathcal{B}_f^*\left(f_{t+1}-\mu_f^*-\Phi_f^* f_t-\Pi_f^* f_t^L\right)\right)^\top S_x^{*,+}\left(\mathcal{B}_f^*\left(f_{t+1}-\mu_f^*-\Phi_f^* f_t-\Pi_f^* f_t^L\right)\right)\\
&= \quad \left(f_{t+1}-\mu_f^*-\Phi_f^* f_t-\Pi_f^* f_t^L\right)^\top S_f^{-1}\left(f_{t+1}-\mu_f^*-\Phi_f^* f_t-\Pi_f^* f_t^L\right)\\
&= \quad \left(f_{t+1}-\mu_{f,t}^*\right)^\top S_f^{-1}\left(f_{t+1}-\mu_{f,t}^*\right)
\end{aligned}$$

where we have used the definition of the generalized inverse above.

Using these expressions, we can derive the log stochastic discount factor

$$\begin{aligned}
\log\mathcal{M}_{t,t+1} \quad &= \quad -\ell_t-\frac{1}{2}\log\det^*\left(2\pi S_x^*\right)-\frac{1}{2}\left(x_{t+1}-\mu_{x,t}^*\right)^\top S_x^{*,+}\left(x_{t+1}-\mu_{x,t}^*\right)\\
&\quad +\frac{1}{2}\log\det^*\left(2\pi S_x\right)+\frac{1}{2}\left(x_{t+1}-\mu_{x,t}\right)^\top S_x^+\left(x_{t+1}-\mu_{x,t}\right)
\end{aligned}$$

When $\beta_f=\beta_f^*$, the pseudo-determinants cancel. This gives

$$\begin{aligned}
\log\mathcal{M}_{t,t+1} \quad &= \quad -\delta_{\ell,0}-\delta_{\ell,x}^\top x_t-\frac{1}{2}\left(f_{t+1}-\mu_{f,t}^*\right)^\top S_f^{-1}\left(f_{t+1}-\mu_{f,t}^*\right)+\frac{1}{2}\left(f_{t+1}-\mu_{f,t}\right)^\top S_f^{-1}\left(f_{t+1}-\mu_{f,t}\right)\\
&= \quad -\delta_{\ell,0}-\delta_{\ell,x}^\top x_t-\frac{1}{2}\left(\mu_{f,t}-\mu_{f,t}^*\right)^\top S_f^{-1}\left(\mu_{f,t}-\mu_{x,t}^*\right)+\mu_{f,t}^\top S_f^{-1}\mu_{x,t}\\
&\quad -\mu_{f,t}^{*,\top}S_f^{-1}\mu_{f,t}-f_{t+1}^\top S_f^{-1}\left(\mu_t-\mu_t^*\right)\\
&= \quad -\delta_{\ell,0}-\delta_{\ell,x}^\top x_t-\frac{1}{2}\left(\mu_{f,t}-\mu_{f,t}^*\right)^\top S_f^{-1}\left(\mu_{f,t}-\mu_{f,t}^*\right)-\tilde{\varepsilon}_{t+1}^\top S_f^{-1}\left(\mu_{f,t}-\mu_{f,t}^*\right)\\
&= \quad -\delta_{\ell,0}-\delta_{\ell,x}^\top x_t-\frac{1}{2}\lambda_t^\top\lambda_t-\lambda_t^\top\varepsilon_{t+1}
\end{aligned}$$

where

$$\begin{aligned}
\lambda_t &= \Sigma_f^{-1}\left(\mu_{f,t}-\mu_{f,t}^*\right) \;=\; \Sigma_f^{-1}\left(\mu_f-\mu_f^*+\left(\Phi_f-\Phi_f^*\right)f_t+\left(\Pi_f-\Pi_f^*\right)f_t^L\right) \\
&= \Sigma_f^{-1}\left(\mu_f-\mu_f^*+\left(\Phi_f-\Phi_f^*\right)f_t+\left(\alpha_f-\alpha_f^*\right)\beta_f^\top f_t^L\right) \;=\; \Sigma_f^{-1}\left(\mu_f-\mu_f^*+\left(\Phi_f-\Phi_f^*\right)f_t+\left(\alpha_f-\alpha_f^*\right)e_t\right)
\end{aligned}$$

This defines the market prices of risk

$$\lambda_0 \;=\; \mu_f-\mu_f^* \qquad \lambda_f \;=\; \Phi_f-\Phi_f^* \qquad \lambda_e \;=\; \alpha_f-\alpha_f^*$$

Appendix D Bond prices

Appendix D.1 U.S. Libor bonds

The price of a 1-period Libor bond is

$$L_t^1 \;=\; \exp\left(\bar{a}_1+\bar{b}_{1,x}^\top x_t\right)$$

where $\bar{a}_1=-\delta_{\ell,0}$ and $\bar{b}_{1,x}=-\delta_{\ell,x}$. The price of an n-period nominal bond is

$$\begin{aligned}
L_t^n &= E_t^*\left[\exp\left(-\ell_t\right)L_{t+1}^{n-1}\right] \;=\; E_t^*\left[\exp\left(-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\bar{a}_{n-1}+\bar{b}_{n-1,x}^\top x_{t+1}\right)\right] \\
&= \exp\left(\bar{a}_{n-1}-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\bar{b}_{n-1,x}^\top\left[\mu_x^*+\Phi_x^* x_t\right]\right)E_t^*\left[\exp\left(\bar{b}_{n-1,x}^\top\Sigma_x\varepsilon_{t+1}\right)\right] \\
&= \exp\left(\bar{a}_{n-1}-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\bar{b}_{n-1,x}^\top\left[\mu_x^*+\Phi_x^* x_t\right]+\frac{1}{2}\bar{b}_{n-1,x}^\top\Sigma_x\Sigma_x^\top\bar{b}_{n-1,x}\right)
\end{aligned}$$

This implies that $L_t^n \;=\; \exp\left(\bar{a}_n+\bar{b}_{n,x}^\top x_t\right)$ where

$$\begin{aligned}
\bar{a}_n &= \bar{a}_{n-1}-\delta_{\ell,0}+\bar{b}_{n-1,x}^\top\mu_x^*+\frac{1}{2}\bar{b}_{n-1,x}^\top\Sigma_x\Sigma_x^\top\bar{b}_{n-1,x} \\
\bar{b}_{n,x} &= \Phi_x^{*\top}\bar{b}_{n-1,x}-\delta_{\ell,x}
\end{aligned}$$

Libor rates are

$$\ell_t^n \;=\; a_n+b_{n,x}^\top x_t$$

where $a_n=-n^{-1}\bar{a}_n$ and $b_{n,x}=-n^{-1}\bar{b}_{n,x}$.

Appendix D.2 Foreign Libor bond prices

The price of an 1-period foreign, nominal Libor bond is

$$\begin{aligned}
\widehat{L}_t^1 &= E_t^*\left[\exp\left(-\ell_t\right)\frac{S_{t+1}}{S_t}\right] \\
&= E_t^*\left[\exp\left(-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\Delta s_{t+1}\right)\right] \;=\; E_t^*\left[\exp\left(\delta_{s,0}-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\delta_{s,x}^\top x_{t+1}\right)\right] \\
&= E_t^*\left[\exp\left(\delta_{s,0}-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\delta_{s,x}^\top\left[\mu_x^*+\Phi_x^* x_t\right]+\delta_{s,x}^\top\Sigma_x\varepsilon_{t+1}\right)\right] \\
&= \exp\left(\delta_{s,0}-\delta_{\ell,0}-\delta_{\ell,x}^\top x_t+\delta_{s,x}^\top\left[\mu_x^*+\Phi_x^* x_t\right]+\frac{1}{2}\delta_{s,x}^\top\Sigma_x\Sigma_x^\top\delta_{s,x}\right)
\end{aligned}$$

This implies that $\widehat{L}_t^1 \;=\; \exp\left(\bar{\bar{a}}_1 + \bar{\bar{b}}_{1,x}^\top x_t\right)$ where

$$\begin{aligned}\bar{\bar{a}}_1 &= \delta_{s,0} - \delta_{\ell,0} + \delta_{s,x}^\top \mu_x^* + \frac{1}{2}\delta_{s,x}^\top \Sigma_x \Sigma_x^\top \delta_{s,x} \\ \bar{\bar{b}}_{1,x} &= \Phi_x^{*\top} \delta_{s,x} - \delta_{\ell,x}\end{aligned}$$

The price of an n-period nominal bond is

$$\begin{aligned}\widehat{L}_t^n &= E_t^*\left[\exp\left(-\ell_t\right)\frac{S_{t+1}}{S_t}\widehat{L}_{t+1}^{n-1}\right] \\ &= \exp\left(\bar{d}_{n-1} + \delta_{s,0} - \delta_{\ell,0} + \left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \left[\mu_x^* + \Phi_x^* x_t\right]\right) E_t^*\left[\exp\left(\left[\left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \Sigma_x\right]\varepsilon_{t+1}\right)\right] \\ &= \exp\left(\bar{\bar{a}}_{n-1} + \delta_{s,0} - \delta_{\ell,0} - \delta_{\ell,x}^\top x_t + \left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \left[\mu_x^* + \Phi_x^* x_t\right]\right) \\ &\quad \exp\left(\frac{1}{2}\left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \Sigma_x \Sigma_x^\top \left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)\right)\end{aligned}$$

This implies that $\widehat{L}_t^n \;=\; \exp\left(\bar{\bar{a}}_n + \bar{\bar{b}}_{n,x}^\top x_t\right)$ where

$$\begin{aligned}\bar{\bar{a}}_n &= \bar{\bar{a}}_{n-1} + \delta_{s,0} - \delta_{\ell,0} + \left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \mu_x^* + \frac{1}{2}\left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right)^\top \Sigma_x \Sigma_x^\top \left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right) \\ \bar{\bar{b}}_{n,x} &= \Phi_x^{*\top}\left(\bar{\bar{b}}_{n-1,x} + \delta_{s,x}\right) - \delta_{\ell,x}\end{aligned}$$

Yields are

$$\widehat{\ell}_t^n \;=\; \widehat{a}_n + \widehat{b}_{n,x}^\top x_t$$

where $\widehat{a}_n = -n^{-1}\bar{\bar{a}}_n$ and $\widehat{b}_{n,x} = -n^{-1}\bar{\bar{b}}_{n,x}$.

Appendix D.2.1 U.S. government bond prices

The price of an 1-period nominal bond is

$$\begin{aligned}Q_t^1 &= E_t^*\left[\exp\left(-\left[\ell_t - c_t\right]\right)\right] \;=\; E_t^*\left[\exp\left(-\delta_{\ell,0} + \delta_{c,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t\right)\right] \\ &= E_t^*\left[\exp\left(-\delta_{\ell,0} + \delta_{c,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t\right)\right] \\ &= \exp\left(-\delta_{\ell,0} + \delta_{c,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t\right)\end{aligned}$$

This implies that $Q_t^1 \;=\; \exp\left(\bar{d}_1 + \bar{h}_{1,x}^\top x_t\right)$ where

$$\begin{aligned}\bar{d}_1 &= -\delta_{\ell,0} + \delta_{c,0} \\ \bar{h}_{1,x} &= -\delta_{\ell,x} + \delta_{c,x}\end{aligned}$$

The price of an n-period nominal bond is

$$\begin{aligned}Q_t^n &= E_t^*\left[\exp\left(-\left[\ell_t - c_t\right]\right)Q_{t+1}^{n-1}\right] \;=\; E_t^*\left[\exp\left(\delta_{c,0} - \delta_{\ell,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t + \bar{d}_{n-1} + \bar{h}_{n-1,x}^\top x_{t+1}\right)\right] \\ &= \exp\left(\bar{d}_{n-1} - \delta_{\ell,0} + \delta_{c,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t + \bar{h}_{n-1,x}^\top\left[\mu_x^* + \Phi_x^* x_t\right]\right) E_t^*\left[\exp\left(\bar{h}_{n-1,x}^\top \Sigma_x \varepsilon_{t+1}\right)\right] \\ &= \exp\left(\bar{d}_{n-1} - \delta_{\ell,0} + \delta_{c,0} - \left(\delta_{\ell,x} - \delta_{c,x}\right)^\top x_t + \bar{h}_{n-1,x}^\top\left[\mu_x^* + \Phi_x^* x_t\right] + \frac{1}{2}\bar{h}_{n-1,x}^\top \Sigma_x \Sigma_x^\top \bar{h}_{n-1,x}\right)\end{aligned}$$

This implies that $Q_t^n = \exp\left(\bar{d}_n + \bar{h}_{n,x}^\top x_t\right)$ where

$$\begin{aligned}
\bar{d}_n &= \bar{d}_{n-1} - \delta_{\ell,0} + \delta_{c,0} + \bar{h}_{n-1,x}^\top \mu_x^* + \frac{1}{2}\bar{h}_{n-1,x}^\top \Sigma_x \Sigma_x^\top \bar{h}_{n-1,x} \\
\bar{h}_{n,x} &= \Phi_x^{*\top} \bar{h}_{n-1,x} - \delta_{\ell,x} + \delta_{c,x}
\end{aligned}$$

Government yields are

$$y_t^n = d_n + h_{n,x}^\top x_t$$

where $d_n = -n^{-1}\bar{d}_n$ and $h_{n,x} = -n^{-1}\bar{h}_{n,x}$.

Appendix D.2.2 Foreign government bond prices

The price of an 1-period nominal bond is

$$\begin{aligned}
\widehat{Q}_t^1 &= E_t^*\left[\exp\left(-\left[\ell_t - \widehat{c}_t\right]\right)\frac{S_{t+1}}{S_t}\right] = E_t^*\left[\exp\left(-\delta_{\ell,0} + \widehat{\delta}_{c,0} - \left(\delta_{\ell,x} - \widehat{\delta}_{c,x}\right)^\top x_t + \Delta s_{t+1}\right)\right] \\
&= \exp\left(\delta_{s,0} - \delta_{\ell,0} + \widehat{\delta}_{c,0} - \left(\delta_{\ell,x} - \widehat{\delta}_{c,x}\right)^\top x_t + \delta_{s,x}^\top\left[\mu_x^* + \Phi_x^* x_t\right] + \frac{1}{2}\delta_{s,x}^\top \Sigma_x \Sigma_x^\top \delta_{s,x}\right)
\end{aligned}$$

This implies that $\widehat{Q}_t^1 = \exp\left(\bar{\widehat{d}}_1 + \bar{\widehat{h}}_{1,x}^\top x_t\right)$ where

$$\begin{aligned}
\bar{\widehat{d}}_1 &= \delta_{s,0} - \delta_{\ell,0} + \hat{\delta}_{c,0} + \delta_{s,x}^\top \mu_x^* + \frac{1}{2}\delta_{s,x}^\top \Sigma_x \Sigma_x^\top \delta_{s,x} \\
\bar{\widehat{h}}_{1,x} &= \Phi_x^{*\top} \delta_{s,x} - \delta_{\ell,x} + \hat{\delta}_{c,x}
\end{aligned}$$

The price of an n-period nominal bond is

$$\begin{aligned}
\widehat{Q}_t^n &= E_t^*\left[\exp\left(-\left[\ell_t - \widehat{c}_t\right]\right)\frac{S_{t+1}}{S_t}\widehat{Q}_{t+1}^{n-1}\right] \\
&= E_t^*\left[\exp\left(\delta_{s,0} + \widehat{\delta}_{c,0} - \delta_{\ell,0} - \left(\delta_{\ell,x} - \widehat{\delta}_{c,x}\right)^\top x_t + \bar{\widehat{d}}_{n-1} + \bar{\widehat{h}}_{n-1,x}^{g,\top} x_{t+1}\right)\right] \\
&= \exp\left(\bar{\widehat{d}}_{n-1} + \delta_{s,0} - \delta_{\ell,0} + \widehat{\delta}_{c,0} - \left(\delta_{\ell,x} - \widehat{\delta}_{c,x}\right)^\top x_t + \left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right)^\top \left[\mu_x^* + \Phi_x^* x_t\right]\right) \\
&\quad \exp\left(\frac{1}{2}\left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right)^\top \Sigma_x \Sigma_x^\top \left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right)\right)
\end{aligned}$$

This implies that $\widehat{Q}_t^n = \exp\left(\bar{\widehat{d}}_n + \bar{\widehat{h}}_{n,x}^\top x_t\right)$ where

$$\begin{aligned}
\bar{\widehat{d}}_n &= \bar{\widehat{d}}_{n-1} + \delta_{s,0} - \delta_{\ell,0} + \widehat{\delta}_{c,0} + \left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right)^\top \mu_x^* + \frac{1}{2}\left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right)^\top \Sigma_x \Sigma_x^\top \left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right) \\
\bar{\widehat{h}}_{n,x} &= \Phi_x^{*\top}\left(\bar{\widehat{h}}_{n-1,x} + \delta_{s,x}\right) - \delta_{\ell,x} + \widehat{\delta}_{c,x}
\end{aligned}$$

Foreign government yields are

$$\widehat{y}_t^n = \widehat{d}_n + \widehat{h}_{n,x}^\top x_t$$

where $\widehat{d}_n = -n^{-1}\bar{\widehat{d}}_n$ and $\widehat{h}_{n,x} = -n^{-1}\bar{\widehat{h}}_{n,x}$.

Appendix E Rotation and Identification

In this appendix, we illustrate how to impose restrictions on the model to allow the state vector to be any linear combination of the observables (macroeconomic variables and yields) chosen by the researcher. We also discuss identification of the model.

Appendix E.1 Rotating the state vector to observables

Define the $d_y \times 1$ vector of observables Y_t as

$$Y_t = \begin{pmatrix} \Delta m_t \\ \Delta_c \ell_t \\ y_t \\ \widehat{y}_t \end{pmatrix}$$

where Δm_t is a vector of stationary macro variables, $\Delta_c \ell_t$ is a vector of Libor rate differences, y_t are U.S. government yields, and $\widehat{y}_t$ are foreign yields. Let W_1 and W_2 denote $d_f \times d_y$ and $d_y - d_f \times d_y$ matrices, that when stacked produce a full rank matrix. These matrices are chosen by the researcher. Using W_1 and W_2, we define two linear combinations of the data

$$\begin{aligned} Y_t^{(1)} &= W_1 Y_t \\ Y_t^{(2)} &= W_2 Y_t \end{aligned}$$

Following the term structure literature, we assume that $Y_t^{(1)}$ is observed without error while $Y_t^{(2)}$ is a vector observed with error. The specific choice of W_1 and W_2 used in the paper are described in Appendix F.2.

We start by re-defining the model in terms of a vector of latent state variables $\tilde{x}_t$ that are an unknown linear combination of the data. The two state vectors x_t and $\tilde{x}_t$ are related to one another via an affine transformation

$$x_t = \Gamma_0 + \Gamma_1 \tilde{x}_t \tag{E.1}$$

For a given W_1, we want to determine how to choose Γ_0 and Γ_1 in order to guarantee the state vector is

$$x_t = \begin{pmatrix} f_t \\ e_t \end{pmatrix} = \begin{pmatrix} W_1 Y_t \\ e_t \end{pmatrix}$$

and that x_t is a cointegrated system as in Appendix C.1. We partition the rotation matrices in blocks as

$$\begin{pmatrix} f_t \\ e_t \end{pmatrix} = \begin{pmatrix} \Gamma_{0,f} \\ \Gamma_{0,e} \end{pmatrix} + \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix} \begin{pmatrix} \tilde{f}_t \\ \tilde{e}_t \end{pmatrix} \tag{E.2}$$

The matrices Γ_{ff}, Γ_{fe} and $\Gamma_{0,f}$ are determined by the choice of W_1. The matrices Γ_{ef}, Γ_{ee} and vector $\Gamma_{0,e}$ have to satisfy internal consistency conditions in order to guarantee that x_t is a cointegrated system.

The risk-adjusted and actual dynamics of the state vector under the latent factor rotation are

$$\begin{aligned} \Delta m_t &= \tilde{\delta}_{m,0} + \tilde{\delta}_{m,x} \tilde{x}_t & \text{(E.3)} \\ \ell_t &= \tilde{\delta}_{\ell,0} + \tilde{\delta}_{\ell,x}^{\top} \tilde{x}_t & \text{(E.4)} \\ \tilde{x}_t &= \tilde{\mu}_x^* + \tilde{\Phi}_x^* \tilde{x}_{t-1} + \tilde{\Sigma}_x \varepsilon_t & \text{(E.5)} \\ \tilde{x}_t &= \tilde{\mu}_x + \tilde{\Phi}_x \tilde{x}_{t-1} + \tilde{\Sigma}_x \varepsilon_t & \text{(E.6)} \end{aligned}$$

We use a tilde $\tilde{\theta}$ on any parameters to distinguish them from the parameters θ of the rotation in terms of observable factors x_t.

According to the model, the observed data Y_t is related to the latent state vector as

$$Y_t = \begin{pmatrix} \tilde{\delta}_{m,0} \\ \tilde{A} - \tilde{\hat{A}} \\ \tilde{D} \\ \tilde{\hat{D}} \end{pmatrix} + \begin{pmatrix} \tilde{\delta}_{m,x} \\ \tilde{B}_x - \tilde{\hat{B}}_x \\ \tilde{H}_x \\ \tilde{\hat{H}}_x \end{pmatrix} \tilde{x}_t = \tilde{M} + \tilde{N}_x \tilde{x}_t$$

where $\tilde{M}$ and $\tilde{N}_x$ collect all the factor loadings. Pre-multiplying by W_1, we find

$$\begin{aligned} Y_t &= \tilde{M} + \tilde{N}_x \tilde{x}_t \\ W_1 Y_t &= W_1 \tilde{M} + W_1 \tilde{N}_x \tilde{x}_t \\ Y_t^{(1)} &= W_1 \tilde{M} + W_1 \tilde{N}_x \Gamma_1^{-1} (x_t - \Gamma_0) \\ Y_t^{(1)} &= W_1 \tilde{M} - W_1 \tilde{N}_x \Gamma_1^{-1} \Gamma_0 + W_1 \tilde{N}_x \Gamma_1^{-1} x_t \end{aligned}$$

In order for $f_t = Y_t^{(1)}$, the rotation requires that two conditions are met

$$\begin{aligned} W_1 \tilde{M} - W_1 \tilde{M}_x \Gamma_1^{-1} \Gamma_0 &= 0 \\ W_1 \tilde{N}_x \Gamma_1^{-1} &= \begin{pmatrix} \mathrm{I} & 0 \end{pmatrix} \end{aligned}$$

We use these conditions to solve for Γ_{ff}, Γ_{fe} and $\Gamma_{0,f}$ in (E.2). We find

$$\begin{aligned} \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \end{pmatrix} &= W_1 \tilde{N}_x \\ \Gamma_{0,f} &= W_1 \tilde{M} \end{aligned}$$

Therefore, the matrices Γ_{ff}, Γ_{fe} and the vector $\Gamma_{0,f}$ are determined by the choice of W_1.

We still need to determine the unknown matrices Γ_{ef}, Γ_{ee} and the vector $\Gamma_{0,e}$ in (E.2). How a researcher must choose these matrices depends on how they parameterize the autocovariance matrix $\tilde{\Phi}_x^*$ and drift $\tilde{\mu}_x^*$ under the latent factor representation.

We parameterize the matrix $\tilde{\Phi}_x^*$ in (E.5) as a matrix of eigenvalues

$$\tilde{\Phi}_x^* = \begin{pmatrix} \Lambda_f^* & \Lambda_{fe}^* \\ \Lambda_{ef}^* & \Lambda_e^* \end{pmatrix}$$

where the blocks are the same dimension as f_t and e_t, respectively. In general, the eigenvalues may be distinct and real, complex, or repeated. In most settings, empirical researchers impose the restrict that the eigenvalues are distinct and real meaning that Λ_f^* and Λ_e^* are diagonal matrices and $\Lambda_{fe}^* = 0$, $\Lambda_{ef}^* = 0$.

Note that under this rotation, the matrix $\tilde{\Phi}_x^*$ implies that the factors $\tilde{x}_t$ are not cointegrated because this matrix does not have the structure of a cointegrated system. The relationship between the autocovariance matrices under the two rotations is

$$\Phi_x^* = \Gamma_1 \tilde{\Phi}_x^* \Gamma_1^{-1}$$

In order for x_t to be cointegrated with cointegrating vector β_f^*, the autocovariance matrix Φ_x^* must have the structure $\Phi_x^* = \mathcal{I} + \mathcal{B}_f^* \mathcal{A}_f^*$, see Appendix C.1. We use this to determine the values of Γ_{ef} and Γ_{ee} that maintain internal consistency in the model. We start by writing

$$
\begin{aligned}
\mathcal{I}+\mathcal{B}_f^*\mathcal{A}_f^* &= \Gamma_1\tilde{\Phi}_x^*\Gamma_1^{-1} \\
\mathcal{I}\Gamma_1+\mathcal{B}_f^*\mathcal{A}_f^*\Gamma_1 &= \Gamma_1\tilde{\Phi}_x^* \\
\begin{pmatrix} 0 & 0 \\ 0 & \mathrm{I} \end{pmatrix}\begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}+\mathcal{B}_f^*\begin{pmatrix} \Phi_f^* & \alpha_f^* \end{pmatrix}\begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix} &= \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}\tilde{\Phi}_x \\
\begin{pmatrix} 0 & 0 \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}+\mathcal{B}_f^*\begin{pmatrix} \Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef} & \Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee} \end{pmatrix} &= \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}\tilde{\Phi}_x \\
\begin{pmatrix} 0 & 0 \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}+\begin{pmatrix} \mathrm{I} \\ \beta_f^{*,\top} \end{pmatrix}\begin{pmatrix} \Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef} & \Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee} \end{pmatrix} &= \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}\begin{pmatrix} \Lambda_f^* & \Lambda_{fe}^* \\ \Lambda_{ef}^* & \Lambda_e^* \end{pmatrix} \\
\begin{pmatrix} 0 & 0 \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix}+\begin{pmatrix} \Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef} & \Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee} \\ \beta_f^{*,\top}\left(\Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef}\right) & \beta_f^{*,\top}\left(\Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee}\right) \end{pmatrix} &= \begin{pmatrix} \Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^* & \Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^* \\ \Gamma_{ef}\Lambda_f^*+\Gamma_{ee}\Lambda_{ef}^* & \Gamma_{ef}\Lambda_{fe}^*+\Gamma_{ee}\Lambda_e^* \end{pmatrix}
\end{aligned}
$$

Next, we guess and verify that Γ_{ef} and Γ_{ee} have the form

$$
\begin{aligned}
\Gamma_{ef} &= \beta_f^{*,\top}J_f \\
\Gamma_{ee} &= \beta_f^{*,\top}J_e
\end{aligned}
$$

where J_f and J_e are to be determined. β_f^* is the cointegrating vector under the risk adjusted distribution. In our case, the cointegrating vectors $\beta_f^*=\beta_f$ are equal under the two distributions and known in advance. In other applications, they may be different under both distributions.

Plugging in these solution gives

$$
\begin{pmatrix} \Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef} & \Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee} \\ \beta_f^{*,\top}\left(J_f+\Phi_f^*\Gamma_{ff}+\alpha_f^*\Gamma_{ef}\right) & \beta_f^{*,\top}\left(J_e+\Phi_f^*\Gamma_{fe}+\alpha_f^*\Gamma_{ee}\right) \end{pmatrix} = \begin{pmatrix} \Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^* & \Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^* \\ \beta_f^{*,\top}\left(J_f\Lambda_f^*+J_e\Lambda_{ef}^*\right) & \beta_f^{*,\top}\left(J_f\Lambda_{fe}^*+J_e\Lambda_e^*\right) \end{pmatrix}
$$

Substitute the top two equations into the bottom two equations to write this system as

$$
\begin{aligned}
\beta_f^{*,\top}\left(J_f+\Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^*\right) &= \beta_f^{*,\top}\left(J_f\Lambda_f^*+J_e\Lambda_{ef}^*\right) \\
\beta_f^{*,\top}\left(J_e+\Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^*\right) &= \beta_f^{*,\top}\left(J_f\Lambda_{fe}^*+J_e\Lambda_e^*\right)
\end{aligned}
$$

In general, we first solve for J_f as a function of J_e.

$$
\begin{aligned}
J_f+\Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^* &= J_f\Lambda_f^*+J_e\Lambda_{ef}^* \\
J_f\left(\mathrm{I}-\Lambda_f^*\right) &= J_e\Lambda_{ef}^*-\Gamma_{ff}\Lambda_f^*-\Gamma_{fe}\Lambda_{ef}^* \\
J_f &= \left[J_e\Lambda_{ef}^*-\Gamma_{ff}\Lambda_f^*-\Gamma_{fe}\Lambda_{ef}^*\right]\left(\mathrm{I}-\Lambda_f^*\right)^{-1}
\end{aligned}
$$

Plugging this into the second equation

$$
\begin{aligned}
J_e+\Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^* &= J_f\Lambda_{fe}^*+J_e\Lambda_e^* \\
J_e+\Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^* &= \left[J_e\Lambda_{ef}^*-\Gamma_{ff}\Lambda_f^*-\Gamma_{fe}\Lambda_{ef}^*\right]\left(\mathrm{I}-\Lambda_f^*\right)^{-1}\Lambda_{fe}^*+J_e\Lambda_e^* \\
J_e\left(\mathrm{I}-\Lambda_e^*-\Lambda_{ef}^*\left(\mathrm{I}-\Lambda_f^*\right)^{-1}\Lambda_{fe}^*\right) &= -\left[\Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^*\right]\left(\mathrm{I}-\Lambda_f^*\right)^{-1}\Lambda_{fe}^*-\Gamma_{ff}\Lambda_{fe}^*-\Gamma_{fe}\Lambda_e^*
\end{aligned}
$$

The solution for J_e is

$$
J_e = -\left[\left(\Gamma_{ff}\Lambda_f^*+\Gamma_{fe}\Lambda_{ef}^*\right)\left(\mathrm{I}-\Lambda_f^*\right)^{-1}\Lambda_{fe}^*+\Gamma_{ff}\Lambda_{fe}^*+\Gamma_{fe}\Lambda_e^*\right]\left(\mathrm{I}-\Lambda_e^*-\Lambda_{ef}^*\left(\mathrm{I}-\Lambda_f^*\right)^{-1}\Lambda_{fe}^*\right)^{-1}
$$

Once J_e is known, we can solve for J_f from the equation above.

In the special case where Φ_x^* is a diagonal matrix with real, distinct eigenvalues, the rotation matrices Γ_{ef} and Γ_{ee} simply to

$$
\begin{aligned}
\Gamma_{ef} &= -\beta_f^{*,\top}\Gamma_{ff}\Lambda_f^*\left(\mathrm{I}-\Lambda_f^*\right)^{-1} \\
\Gamma_{ee} &= -\beta_f^{*,\top}\Gamma_{fe}\Lambda_e^*\left(\mathrm{I}-\Lambda_e^*\right)^{-1}
\end{aligned}
$$

The state vector is also shifted by Γ_0 in (E.1). We need to determine the value of $\Gamma_{0,e}$ in (E.2) such that after rotating the latent factor $\tilde{e}_t$ becomes e_t. Let $\Upsilon^* = \mathrm{I} - \Phi_x^*$. The relationship between the two rotations is

$$\mu_x^* = (\mathrm{I} - \Phi_x^*)\,\Gamma_0 + \Gamma_1 \tilde{\mu}_x^*$$

In order for x_t to be cointegrated, the drift must have the structure $\mu_x^* = \mathcal{C} + \mathcal{B}_f^* \mu_f^*$, see Appendix C.1. We use this to solve for the value of $\Gamma_{0,e}$.

$$\begin{pmatrix} 0 \\ \beta_f^{*,\top}\delta_0 \end{pmatrix} + \begin{pmatrix} \mu_f^* \\ \beta_f^{*,\top}\mu_f^* \end{pmatrix} = \begin{pmatrix} \Upsilon_{ff}^* & \Upsilon_{fe}^* \\ \Upsilon_{ef}^* & \Upsilon_{ee}^* \end{pmatrix} \begin{pmatrix} \Gamma_{0,f} \\ \Gamma_{0,e} \end{pmatrix} + \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \\ \Gamma_{ef} & \Gamma_{ee} \end{pmatrix} \left[\begin{pmatrix} 0 \\ \beta_f^{*,\top}\tilde{\delta}_0 \end{pmatrix} + \begin{pmatrix} \tilde{\mu}_f^* \\ \tilde{\mu}_e^* \end{pmatrix} \right]$$

Plugging the second equation into the first, this implies that

$$\begin{aligned}
\beta_f^{*,\top}\delta_0 + \beta_f^{*,\top}\mu_f^* &= \Upsilon_{ef}^*\Gamma_{0,f} + \Upsilon_{ee}^*\Gamma_{0,e} + \Gamma_{ef}\tilde{\mu}_f^* + \Gamma_{ee}\tilde{\mu}_e^* + \Gamma_{ee}\beta_f^{\top}\tilde{\delta}_0 \\
\beta_f^{*,\top}\left(\Upsilon_{ff}^*\Gamma_{0,f} + \Upsilon_{fe}^*\Gamma_{0,e} + \Gamma_{ff}\tilde{\mu}_f^* + \Gamma_{fe}\tilde{\mu}_e^*\right) &= -\beta_f^{*,\top}\delta_0 + \Upsilon_{ef}^*\Gamma_{0,f} + \Upsilon_{ee}^*\Gamma_{0,e} + \Gamma_{ef}\tilde{\mu}_f^* + \Gamma_{ee}\tilde{\mu}_e^* + \Gamma_{ee}\beta_f^{*,\top}\tilde{\delta}_0 \\
\left(\beta_f^{*,\top}\Upsilon_{fe}^* - \Upsilon_{ee}^*\right)\Gamma_{0,e} &= -\beta_f^{*,\top}\delta_0 + \Upsilon_{ef}^*\Gamma_{0,f} + \Gamma_{ef}\tilde{\mu}_f^* + \Gamma_{ee}\tilde{\mu}_e^* \\
&\quad -\beta_f^{*,\top}\left(\Upsilon_{ff}^*\Gamma_{0,f} + \Gamma_{ff}\tilde{\mu}_f^* + \Gamma_{fe}\tilde{\mu}_e^*\right) + \Gamma_{ee}\beta_f^{*,\top}\tilde{\delta}_0
\end{aligned}$$

The solution is

$$\begin{aligned}
\Gamma_{0,e} &= \left(\beta_f^{*,\top}\Upsilon_{fe}^* - \Upsilon_{ee}^*\right)^{-1}\Big[\Gamma_{ee}\beta_f^{*,\top}\tilde{\delta}_0 - \beta_f^{*,\top}\delta_0 + \left(\Upsilon_{ef}^* - \beta_f^{*,\top}\Upsilon_{ff}^*\right)\Gamma_{0,f} \\
&\quad + \left(\Gamma_{ef} - \beta_f^{*,\top}\Gamma_{ff}\right)\tilde{\mu}_f^* + \left(\Gamma_{ee} - \beta_f^{*,\top}\Gamma_{fe}\right)\tilde{\mu}_e^*\Big]
\end{aligned}$$

Again, in our case, the vector $\beta_f^* = \beta_f$ are equal and known a priori.

Appendix E.2 Identification

In this section, we discuss how to impose identifying restrictions that are commonly used in the term structure literature. Then, we discuss some specific details used in our estimation.

Identification.

- The parameters of the VECM given by $\mu_f, \Phi_f, \alpha_f, \Sigma_f$ using the observables rotation. They are unrestricted.
- The matrix $\tilde{\Phi}_x^*$ is diagonal with eigenvalues along its diagonal. We assume these eigenvalues are all real and ordered in ascending order.
- The loadings on the short rate are restricted to be a vector of ones $\tilde{\delta}_{\ell,x} = \iota$. The remaining loadings $\tilde{\delta}_{s,x}, \tilde{\delta}_{\pi,x}, \tilde{\delta}_{c,x}$ and $\tilde{\delta}_{\widehat{c},x}$ are unrestricted.
- The vector

$$\tilde{\delta}_0 = \begin{pmatrix} \tilde{\delta}_{s,0} \\ \tilde{\delta}_{\pi,0} \\ \tilde{\delta}_{c,0} \\ \tilde{\delta}_{\widehat{c},0} \\ \tilde{\delta}_{\ell,0} \end{pmatrix}$$

  can be freely estimated.

- The vector $\tilde{\mu}_x^*$ is

$$\tilde{\mu}_x^* = \begin{pmatrix} \tilde{\mu}_f^* \\ \tilde{\mu}_e^* \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{\mu}_e^* \end{pmatrix}$$

where $\tilde{\mu}_f^* = 0$. The value of $\tilde{\mu}_e^*$ must be restricted. In our setting, it is required to be

$$\tilde{\mu}_e^* = (I - \Lambda_e^*)\left(\tilde{\delta}_{s,e} - \tilde{\delta}_{\pi,e}\right)^{-1}\left(\tilde{\delta}_{s,0} - \tilde{\delta}_{\pi,0}\right)$$

The reason the restriction on the drift of $\tilde{\mu}_e^*$ is required is explained in Appendix E.3.

Appendix E.3 Restriction on the drift $\tilde{\mu}_e^*$

Let $\bar{\mu}_x$ and $\bar{\mu}_x^*$ denote the unconditional mean of x_t under the physical and risk-adjusted probabilities. Cointegration imposes cross-equation restrictions on the long-run mean of x_t. Specifically, the cointegrating vector times the unconditional mean of the states f_t must be zero

$$\begin{aligned} \beta_f' S_f \bar{\mu}_x &= 0 \\ \beta_f' S_f \bar{\mu}_x^* &= 0 \end{aligned}$$

where S_f is a selection matrix defined such that $f_t = S_f x_t$. In our setting with $\beta_f^\top = \left(\beta_m^\top \; 0_{1\times d_g}\right)$ and $\beta_m = (1 \;\; -1)^\top$, this restriction implies that the depreciation rate Δs_t and inflation rate differential $\Delta_c \pi_t$ must have the same unconditional mean.

$$\begin{aligned} \bar{\mu}_s &= \bar{\mu}_\pi \\ \bar{\mu}_s^* &= \bar{\mu}_\pi^* \end{aligned}$$

This imposes an implicit restriction on the risk neutral drift of the last factor $\tilde{e}_t$ under the latent factor representation $\tilde{\mu}_e^*$.

To see this, the relationship between the unconditional means across the two rotations is

$$\bar{\mu}_x^* = \Gamma_0 + \Gamma_1 \bar{\tilde{\mu}}_x^*$$

Multiplying both sides by $\beta_f' S_f$ we get

$$\begin{aligned} \beta_f' S_f \bar{\mu}_x^* &= \beta_f' S_f \Gamma_0 + \beta_f' S_f \Gamma_1 \bar{\tilde{\mu}}_x^* \\ 0 &= \beta_f' S_f \Gamma_0 + \beta_f' S_f \Gamma_1 \bar{\tilde{\mu}}_x^* \\ 0 &= \beta_f' \Gamma_{0,f} + \beta_f' \begin{pmatrix} \Gamma_{ff} & \Gamma_{fe} \end{pmatrix} \bar{\tilde{\mu}}_x^* \\ 0 &= \beta_f' \Gamma_{0,f} + \beta_f' \Gamma_{ff} \bar{\tilde{\mu}}_f^* + \beta_f' \Gamma_{fe} \bar{\tilde{\mu}}_e^* \end{aligned}$$

The solution for the mean is

$$\bar{\tilde{\mu}}_e^* = -\left(\beta_f' \Gamma_{fe}\right)^{-1}\left(\beta_f' \Gamma_{0,f} + \beta_f' \Gamma_{ff} \bar{\tilde{\mu}}_f^*\right)$$

We can also write this in terms of the drift as

$$\tilde{\mu}_e^* = -\left(\mathrm{I} - \Lambda_e^*\right)\left(\beta_f' \Gamma_{fe}\right)^{-1}\left(\beta_f' \Gamma_{0,f} + \beta_f' \Gamma_{ff} \bar{\tilde{\mu}}_f^*\right)$$

This condition must be imposed during estimation. In our setting, we know the value of β_f together with the identifying restrictions

$$\begin{aligned} \bar{\tilde{\mu}}_f^* &= 0 \\ \beta_f' \Gamma_{0,f} &= \tilde{\delta}_{s,0} - \tilde{\delta}_{\pi,0} \\ \beta_f' \Gamma_{fe} &= \tilde{\delta}_{s,e} - \tilde{\delta}_{\pi,e} \end{aligned}$$

Plugging these values, we find

$$\tilde{\mu}_e^* = \left(\tilde{\delta}_{s,e} - \zeta_e^*\right)\left(1 - \tilde{\delta}_{p,e}\right)^{-1}\left(\tilde{\delta}_{s,0} - \tilde{\delta}_{\pi,0}\right)$$

Appendix E.4 Parameterization

In this section, we describe how we parameterize the drifts/unconditional means of the model during estimation. For the physical dynamics, we parameterize the model in terms of the unconditional means $\bar{\mu}_x$ instead of the drift μ_f. Instead or parameterizing the model in terms of $\tilde{\delta}_0$ under the latent factor rotation, we prefer to estimate the unconditional means μ_x^* under the observable rotation.

The state vector f_t has dimension $d_f \times 1$. There are d_f free parameters in the vector $\bar{\mu}_x$. As discussed above, cointegration imposes long-run restrictions across those variables such that $\beta_f' S_f \bar{\mu}_x = 0$. In our model, this implies that the unconditional means of the inflation differential and depreciation rate are equal $\bar{\mu}_s = \bar{\mu}_\pi$. Imposing this restriction allows us to estimate the remaining unconditional means of f_t as well as the unconditional mean of the real exchange rate e_t. We write this as

$$\bar{\mu}_x = S_{\bar{\mu}} \bar{\mu}_{x,u}$$

where $S_{\bar{\mu}}$ is a selection matrix that imposes the cointegration restriction $\bar{\mu}_s = \bar{\mu}_\pi$. The vector $\bar{\mu}_{x,u}$ has dimension $d_f \times 1$ and contains the unrestricted means.

From the latent factor rotation, the vector $\tilde{\delta}_0 = \left(\tilde{\delta}_{s,0}, \tilde{\delta}_{\pi,0}, \tilde{\delta}_{c,0}, \tilde{\delta}_{\widehat{c},0}, \tilde{\delta}_{\ell,0}\right)$ contains 5 parameters that are identifiable and there are no free parameters in $\tilde{\mu}_f^*$. In practice, we could estimate these 5 free parameters directly under this rotation. We prefer to parameterize these 5 free parameters in terms of unrestricted values under the observables rotation x_t. The reason is that it is easier to place a prior distribution over $\bar{\mu}_x^*$ in the observables rotation as it is more meaningful. Doing this requires some understanding of the model and how the rotation to observables changes the identifiable parameters that enter $\tilde{\delta}_0$ to free parameters that enter δ_0 and $\bar{\mu}_x^*$.

To start, we know that under our rotation $\delta_{s,0} = \delta_{\pi,0} = \delta_{c,0} = \delta_{\ell,0} = 0$ because all these factors are observable in the state vector f_t. We also know that because the inflation differential and depreciation rate are cointegrated the unconditional means are equal

$$\bar{\mu}_s^* = \bar{\mu}_\pi^*. \tag{E.7}$$

In our setting, $\delta_{\widehat{c},0}$ is still a free parameter because the foreign spread $\widehat{c}_t$ is not observable and does not enter f_t. There are four remaining free parameters in the unconditional mean $\bar{\mu}_x^*$ even though it has dimension $d_x = 9$. These parameters are $(\bar{\mu}_s^*, \bar{\mu}_e^*, \bar{\mu}_c^*, \bar{\mu}_\ell^*)$ which are the unconditional means of $(\Delta s_t, e_t, c_t, \ell_t)$, respectively. The remaining parameters in $\bar{\mu}_x^*$ are not free and are deterministic functions of the estimable parameters $\alpha = (\bar{\mu}_s^*, \bar{\mu}_e^*, \bar{\mu}_c^*, \bar{\mu}_\ell^*, \delta_{\widehat{c},0})$. It turns out that we can solve for these unknown values from knowledge that linear combinations of the loadings M must be zero $W_1 M = 0$ and the unconditional mean restriction in (E.7). First, note that we can always write the loadings as a linear function their means

$$M = M_c + M_{\bar{\mu},r} \bar{\mu}_{x,r}^* + M_{\bar{\mu},u} \bar{\mu}_{x,u}^*$$

We can therefore solve for $\bar{\mu}_{x,r}^*$ as a function of other parameters of the model as

$$\begin{aligned}
\begin{pmatrix} W_1 M \\ \beta_f^\top S_f \bar{\mu}_x^* \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} W_1 M_c + W_1 M_{\bar{\mu},r} \bar{\mu}_{x,r}^* + W_1 M_{\bar{\mu},u} \bar{\mu}_{x,u}^* \\ \beta_f^\top S_{f,r} \bar{\mu}_{x,r}^* + \beta_f^\top S_{f,u} \bar{\mu}_{x,u}^* \end{pmatrix} &= 0 \\
\begin{pmatrix} W_1 M_c \\ 0 \end{pmatrix} + \begin{pmatrix} W_1 M_{\bar{\mu},u} \\ \beta_f^\top S_{f,r} \end{pmatrix} \bar{\mu}_{x,r}^* + \begin{pmatrix} W_1 M_{\bar{\mu},u} \\ \beta_f^\top S_{f,u} \end{pmatrix} \alpha &= 0
\end{aligned}$$

We can solve this system of equations for the restricted parameters $\bar{\mu}_{x,r}^*$ as

$$\bar{\mu}_{x,r}^* = -\left(S_r \begin{pmatrix} W_1 M_{\bar{\mu},r} \\ \beta_f^\top S_{f,r} \end{pmatrix}\right)^{-1} S_r \begin{pmatrix} W_1 M_c \\ 0 \end{pmatrix} - \left(S_r \begin{pmatrix} W_1 M_{\bar{\mu},r} \\ \beta_f^\top S_{f,r} \end{pmatrix}\right)^{-1} S_r \begin{pmatrix} W_1 M_{\bar{\mu},u} \\ \beta_f^\top S_{f,u} \end{pmatrix} \alpha$$

Appendix F Estimation

Appendix F.1 Prior distributions

- Let $S_y = \Sigma_y \Sigma_y^\top$ with dimension $d_{y_2} \times d_{y_2}$. Note that $Y_t^{(2)}$ has dimension $d_{y_2} \times 1$. We assume S_y has an inverse Wishart distribution $S_y \sim \text{Inv-W}\left(\underline{\Omega}_y, \underline{\nu}_y\right)$ with mean $E[S_y] = \Omega_y (\nu - d_{y_2} - 1)^{-1}$. We set $\underline{\nu}_y = d_{y_2} + 3$ and $\underline{\Omega}_y = I_{d_{y_2}} (\nu_y - d_{y_2} - 1) \times 10^{-8}$.
- Let $S_f = \Sigma_f \Sigma_f^\top$ with dimension $d_f \times d_f$. We assume S_f has an inverse Wishart distribution $S_f \sim \text{Inv-W}\left(\underline{\Omega}_f, \underline{\nu}_f\right)$. We set $\underline{\nu}_f = d_f + 3$ and $\underline{\Omega} = \underline{S}_f (\nu_f - d_f - 1)$. The matrix $\underline{S}_f$ is diagonal with blocks $\underline{S}_s, \underline{S}_\pi, \underline{S}_g$ whose dimensions are the same as $\Delta s_t, \Delta_c \pi_t$ and g_t. The scale of the depreciation rate is significantly larger than inflation and yields. We set $\underline{S}_s = \mathrm{I} \times 10^{-3}$, $\underline{S}_\pi = \mathrm{I} \times 10^{-5}$, and $\underline{S}_g = \mathrm{I} \times 10^{-7}$.
- As discussed in Appendix E.4, we estimate the unconditional means $(\bar{\mu}_x, \bar{\mu}_x^*)$ directly instead of the drifts $\left(\mu_f, \mu_f^*\right)$. In a VECM, there are d_f free parameters in $\bar{\mu}_x$; the same as the dimension of f_t. We assume each free entry in $\bar{\mu}_x$ is a-priori independent.

  In our setting, the factors x_t are observable. First, we calculate the unconditional sample mean of the factors $\hat{\bar{\mu}}_x$. Our prior for each element of $\bar{\mu}_x$ is a normal distribution centered at the sample mean. Then, we choose the variance of this distribution to be large enough to cover the support of the data. Let $\kappa = 1/1200^2$. Our priors are

  - inflation rate differential: $\bar{\mu}_\pi \sim \mathrm{N}\left(\hat{\bar{\mu}}_\pi, 25\kappa\right)$
  - Libor rate: $\bar{\mu}_\ell \sim \mathrm{N}\left(\hat{\bar{\mu}}_\ell, \kappa\right)$
  - term spread: $\bar{\mu}_{y^{120,12}} \sim \mathrm{N}\left(\hat{\bar{\mu}}_{y^{120,12}}, 0.5\kappa\right)$
  - Ted spread: $\bar{\mu}_c \sim \mathrm{N}\left(\hat{\bar{\mu}}_c, 0.25\kappa\right)$
  - Interest rate differential: $\bar{\mu}_{\Delta_c \ell} \sim \mathrm{N}\left(\hat{\bar{\mu}}_{\Delta_c \ell}, 0.5\kappa\right)$
  - term spread differential: $\bar{\mu}_{\Delta_c y^{120,12}} \sim \mathrm{N}\left(\hat{\bar{\mu}}_{\Delta_c y^{120,12}}, 0.5\kappa\right)$
  - Libor slope differential: $\bar{\mu}_{\Delta_c \ell^{6,1}} \sim \mathrm{N}\left(\hat{\bar{\mu}}_{\ell^{6,1}}, 0.25\kappa\right)$
  - real exchange rate: $\bar{\mu}_e \sim \mathrm{N}\left(\hat{\bar{\mu}}_e, 5000\kappa\right)$

  No arbitrage imposes restrictions on $\bar{\mu}_x^*$, allowing for only 4 free parameters. We model these using a conditional prior distribution as

  - depreciation rate: $\bar{\mu}_s^* \sim \mathrm{N}\left(\bar{\mu}_s, 50\kappa\right)$
  - Libor rate: $\bar{\mu}_\ell^* \sim \mathrm{N}\left(\bar{\mu}_\ell + 15, 50\kappa\right)$
  - Ted spread: $\bar{\mu}_c^* \sim \mathrm{N}\left(\bar{\mu}_c, 50\kappa\right)$
  - real exchange rate: $\bar{\mu}_e^* \sim \mathrm{N}\left(\bar{\mu}_e, 25000\kappa\right)$

  Note that our conditional prior distribution implicitly restricts the magnitudes of the market prices of risk λ_μ.
- The eigenvalues Λ_x^* of Φ_x^* are assumed to be real and ordered. Let $a_1 = -1$ and $b = 1$. We parameterize them as $\Lambda_{x,1}^* = a_1 + (b - a_1)U_1$ and $\Lambda_{x,j}^* = a_{j-1} + (b - a_{j-1})U_j$ for $j = 2, \ldots, d_x$. This transformation ensures that they are increasing and contained in the interval $[-1, 1]$. We then place priors on the eigenvalues $\Lambda_{x,j}^*$ via $U_j \sim \mathrm{Beta}(10, 10)$.
- We place a prior on the free parameters of the factor loadings $\tilde{\delta}_{s,x}$, $\tilde{\delta}_{\pi,x}$, $\tilde{\delta}_{c,x}$, $\tilde{\delta}_{\widehat{c},x}$. Our identifying restriction is that $\tilde{\delta}_{\ell,x} = \iota$. We assume that each free entry is independent and distributed as $\tilde{\delta}_{j,x} \sim \mathrm{N}(0, 10000)$. This implicitly places a prior distribution over the free parameters in Φ_f^* and α_f^* .
- We use a conditional prior distribution $p\left(\Phi_f | \Phi_f^*\right)$ where $\mathrm{vec}\left(\Phi_f\right) \sim \mathrm{N}\left(\mathrm{vec}\left(\Phi_f^*\right), V_{\phi^*}\right)$. The covariance matrix V_{ϕ^*} then measures the magnitudes of the market prices of risk.
- We use a joint prior distribution over the speed of adjustment parameters $p\left(\alpha_f, \alpha_f^*\right) = p\left(\alpha_f | \alpha_f^*\right) p\left(\alpha_f^*\right)$.

Appendix F.2 Observables

We stack the U.S. and foreign nominal yields of different maturities into vectors $y_t = \left(y_t^3, \ldots, y_t^{120}\right)$ and $\widehat{y}_t = \left(\widehat{y}_t^3, \ldots, \widehat{y}_t^{120}\right)$ as well as their bond loadings, e.g. $D = (d_3, \ldots, d_{120})^\top$ and $H_x = (h_{3,x}, \ldots, h_{120,x})^\top$. We do the same for the observed Libor rate differentials $\Delta_c \ell_t = \left(\Delta_c \ell_t^1, \ldots, \Delta_c \ell_t^{12}\right)$ and their loadings $A = (a_1, \ldots, a_{12})^\top$ and $B_x = (b_{1,x}, \ldots, b_{12,x})^\top$. In practice, we also observe the one month U.S. Ted spread c_t^1 and the one month U.S. Libor rate ℓ_t^1.

The system of observation equations used in the model are

$$
\begin{aligned}
m_t &= m_{t-1} + \delta_{m,0} + \delta_{m,x} x_t \\
\ell_t^1 &= a_1 + b_{1,x}^\top x_t \\
c_t^1 &= (a_1 - d_1) + (b_{1,x} - h_{1,x})^\top x_t \\
\Delta_c \ell_t &= \left(A - \widehat{A}\right) + \left(B_x - \widehat{B}_x\right) x_t \\
y_t &= D + H_x x_t \\
\widehat{y}_t &= \widehat{D} + \widehat{H}_x x_t
\end{aligned}
$$

We define the overall vector of observables as

$$
Y_t = \begin{pmatrix} \Delta m_t \\ \ell_t^1 \\ c_t^1 \\ \Delta_c \ell_t \\ y_t \\ \widehat{y}_t \end{pmatrix}
$$

We choose W_1 so that the state vector f_t is

$$
f_t = \begin{pmatrix} \Delta s_t \\ \Delta_c \pi_t \\ \ell_t^1 \\ y_t^{120,12} \\ c_t^1 \\ \Delta_c \ell_t \\ \Delta_c y_t^{120,12} \\ \Delta_c \ell_t^{12,1} \end{pmatrix}
$$

The matrix W_2 is a selection matrix full of zeros and ones that selects out of Y_t the elements that are not used in f_t. Specifically, W_2 is defined such that $Y_t^{(2)}$ includes the 3, 12, 24, 36, 48, 60, and 84 month U.S. yields, the 3, 24, 36, 48, 60, 84, 120 month foreign yields, and a linear combination of Libor rates $\ell_t^1 - \ell^3 + \widehat{\ell}^3$ and $\ell_t^1 - \ell^6 + \widehat{\ell}^6$. The last two linear combinations of Libor rates where chosen so that they have the same magnitude and sign as foreign government yields.

Appendix F.3 Log-likelihood function

The log-likelihood function is

$$
\mathcal{L} = \log p\left(Y_1, \ldots, Y_T | \theta\right) = \sum_{t=1}^{T} \log p\left(f_t | f_{t-1}, e_{t-1}; \theta\right) + \sum_{t=1}^{T} \log p\left(Y_t^{(2)} | x_t; \theta\right)
$$

where f_0 and e_0 are assumed to be known. The density $p(f_t|f_{t-1}, e_{t-1}; \theta)$ is determined by the VECM dynamics of the factors f_t while the second term comes from the linear combination of yields observed with error

$$Y_t^{(2)} = M^{(2)} + N_x^{(2)} x_t + \Sigma_y \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \mathrm{I}),$$

where $M^{(2)} = W_1 M$ and $N_x^{(2)} = W_2 N_x$ and

$$\begin{aligned} M &= \tilde{M} - \tilde{N}_x \Gamma_1^{-1} \Gamma_0, \\ N_x &= \tilde{N}_x \Gamma_1^{-1}. \end{aligned}$$

This likelihood function assumes that there are no missing values in either $Y_t^{(1)}$ or $Y_t^{(2)}$. In practice, this is not the case. We impute these missing values during the MCMC algorithm using the Kalman filter.

Appendix F.4 Estimation

Let θ denote all the parameters of the model and define $f_{1:T} = (f_1, \ldots, f_T)$ and $Y_{1:T} = (Y_1, \ldots, Y_T)$. In practice, some data points are missing which implies that some of the factors f_t are missing. We use $Y_{1:T}^o$ and $Y_{1:T}^m$ to denote the observed and missing data, respectively. The joint posterior distribution over the parameters and missing data is given by

$$p(\theta, Y_{1:T}^m | Y_{1:T}^o) \propto p(Y_{1:T}^o | \theta) \, p(\theta),$$

where $p(Y_{1:T}^o|\theta)$ is the likelihood and $p(\theta)$ is the prior distribution. We use Markov-chain Monte Carlo to draw from the posterior.

Appendix F.4.1 MCMC algorithm

We provide a brief description of the MCMC algorithm. Let $S_y = \Sigma_y \Sigma_y'$ and $S_f = \Sigma_f \Sigma_f'$ denote the covariance matrices. We use a Gibbs sampler that iterates between drawing from each of the full conditional distributions.

- Place the model in linear, Gaussian state space form as described in Appendix F.5. Draw the missing data and unconditional means $(Y_{1:T}^m, \bar{\mu}_x, \bar{\mu}_x^*)$ from their full conditional distribution using the Kalman filter and simulation smoothing algorithm. Given the full data $Y_t^{o,m} = (Y_t^o, Y_t^m)$, we can recalculate the factors $f_t = W_1 Y_t^{o,m}$.
- Let $\bar{f}_t = f_t - \bar{\mu}_f$ and $\bar{e}_t = e_t - \bar{\mu}_e$ denote the demeaned factors. We draw the free elements of Φ_f, α_f from their full conditional distribution using standard results for Bayesian multiple regression. We write the VECM as a regression model

  $$\bar{f}_t = X_t \gamma + \Sigma_f \varepsilon_t$$

  where $\gamma = (\mathrm{vec}(\Phi_f) \ \ \alpha_f)$ and the regressors X_t contain lagged values of $\bar{f}_{t-1}$ and $\bar{e}_{t-1}$. Draws from this model are standard.
- Draw the free elements of S_f from their full conditional using a random-walk Metropolis algorithm. In this step, we avoid conditioning on the parameters S_y, Φ_f, α_f by analytically integrating these parameters out of the likelihood.
- Draw the eigenvalues Λ_x^* from their full conditional using random-walk Metropolis. To avoid conditioning on S_y, Φ_f, α_f, we draw from the marginal distribution that analytically integrates these values out of the likelihood.
- Draw the elements of $\tilde{\delta}_{s,x}, \tilde{\delta}_{\pi,x}, \tilde{\delta}_{c,x}$ and $\tilde{\delta}_{\widehat{c},x}$ from their full conditional using random-walk Metropolis. To avoid conditioning on S_y, Φ_f, α_f, we draw from the marginal distribution that analytically integrates these values out of the likelihood.
- The full conditional posterior of S_y is an inverse Wishart distribution $S_y \sim \text{Inv-Wish}\left(\bar{\nu}, \bar{\Omega}\right)$ where $\bar{\nu} = \underline{\nu} + T$ and $\bar{\Omega} = \underline{\Omega} + \sum_{t=1}^{T} \eta_t \eta_t^\top$.

Appendix F.5 Imputing missing values

In our data set, some of the macroeconomic variables and yields contain missing values. A missing value of a macroeconomic variable in levels m_t implies two missing values in first differences Δm_t. We therefore formulate the state space model in levels m_t and impute the missing values during estimation under a missing at random assumption. We use Y_t^L to denote the vector of observables that contain the levels of the macro variables and yields

$$Y_t^L = \begin{pmatrix} m_t \\ y_t \end{pmatrix} \qquad Y_t = \begin{pmatrix} \Delta m_t \\ y_t \end{pmatrix}$$

while Y_t contains the first differences of the macro variables.

Given that $f_t = Y_t^{(1)}$, we can write the model in VAR form as

$$\begin{aligned} \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix} &= \begin{pmatrix} \mu_f \\ M^{(2)} + N_x^{(2)} \mathcal{B}_f \mu_f \end{pmatrix} + \begin{pmatrix} \Phi_f & 0 \\ R^{(2)} \Phi_f & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1}^{(1)} \\ Y_{t-1}^{(2)} \end{pmatrix} \\ &+ \begin{pmatrix} \alpha_f \\ N_e^{(2)} + R^{(2)} \alpha_f \end{pmatrix} e_{t-1} + \begin{pmatrix} \Sigma_f & 0 \\ R^{(2)} \Sigma_f & \Sigma_y \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \end{aligned}$$

where $R^{(2)} = N_f^{(2)} + N_e^{(2)} \beta_f^\top$. We use $N_f^{(2)}$ to denote the columns of $N_x^{(2)}$ associated with the factors f_t. A similar definition applies to $N_e^{(2)}$ and e_t. Next we translate this system back into Y_t using the fact that

$$Y_t = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix}$$

to get

$$\begin{aligned} Y_t &= \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_f \\ M^{(2)} + N_x^{(2)} \mathcal{B}_f \mu_f \end{pmatrix} + \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \Phi_f & 0 \\ R^{(2)} \Phi_f & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} Y_{t-1} \\ &+ \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \alpha_f \\ N_e^{(2)} + R^{(2)} \alpha_f \end{pmatrix} e_{t-1} + \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_f & 0 \\ R^{(2)} \Sigma_f & \Sigma_y \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \end{aligned}$$

This structure implies that Y_t is a reduced-rank VECM of the form

$$Y_t = \mu_Y + \Phi_Y Y_{t-1} + \alpha_Y e_{t-1} + \Sigma_Y \varepsilon_{Y,t} \qquad \varepsilon_{Y,t} \sim \mathrm{N}(0, I)$$

where

$$\mu_Y = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_f \\ M^{(2)} + N_x^{(2)} \mathcal{B}_f \mu_f \end{pmatrix} \qquad \Phi_Y = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \Phi_f & 0 \\ R^{(2)} \Phi_f & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$$

$$\alpha_Y = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \alpha_f \\ N_e^{(2)} + R^{(2)} \alpha_f \end{pmatrix} \qquad \Sigma_Y = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_f & 0 \\ R^{(2)} \Sigma_f & \Sigma_y \end{pmatrix} \qquad \varepsilon_{Y,t} = \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

Appendix F.5.1 State space form

We place this model in the following linear, Gaussian state space form

$$Y_t^L = Z\alpha_t + d + u_t \qquad u_t \sim \mathrm{N}(0, H), \tag{F.1}$$

$$\alpha_{t+1} = T\alpha_t + c + Rv_t \qquad v_t \sim \mathrm{N}(0, Q). \tag{F.2}$$

where the initial condition is $\alpha_1 \sim \mathrm{N}\left(a_{1|0}, P_{1|0}\right)$.

Let $\bar{\mu} = \left(\bar{\mu}_{x,u}^{\top}\ \bar{\mu}_{x,u}^{*,\top}\ \delta_{\widehat{c},0}\right)^{\top}$ denote the vector of unrestricted unconditional means that enter $\bar{\mu}_x$ and $\bar{\mu}_x^*$ plus the intercept $\delta_{\widehat{c},0}$. The vector of intercepts μ_Y can be written as a linear function of the unconditional means

$$\mu_Y \quad = \quad S_{\mu,0} + S_{\mu,1}\bar{\mu}$$

We draw unconditional means jointly with the missing data by including them in the state vector. By definition $m_t = m_{t-1} + S_m Y_t$ and $y_t = S_y Y_t$ where S_m and S_y are selection matrices. We define the system matrices from (F.1)-(F.2) as

$$d \;=\; 0 \qquad Z \;=\; \begin{pmatrix} \mathrm{I} & S_m & 0 \\ 0 & S_y & 0 \end{pmatrix} \qquad H = 0 \qquad Q = \Sigma_Y \Sigma_Y^{\top}$$

$$\alpha_t \;=\; \begin{pmatrix} m_{t-1} \\ Y_t \\ \bar{\mu} \end{pmatrix} \qquad T \;=\; \begin{pmatrix} \mathrm{I} & \delta_{m,f} & 0 \\ \alpha_Y \beta_m^{\top} & \Phi_Y & S_{\mu,1} \\ 0 & 0 & \mathrm{I} \end{pmatrix} \qquad c \;=\; \begin{pmatrix} \delta_{m,0} \\ S_{\mu,0} \\ 0 \end{pmatrix} \qquad R \;=\; \begin{pmatrix} 0 \\ \mathrm{I} \\ 0 \end{pmatrix}$$

$$a_{1|0} \;=\; \begin{pmatrix} m_0 \\ \alpha_Y \beta_m^{\top} m_0 + S_{\mu,1}\bar{m}_{\mu} \\ \bar{m}_{\mu} \end{pmatrix} \qquad P_{1|0} \;=\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Sigma_Y \Sigma_Y^{\top} + S_{\mu,1} V_{\mu} S_{\mu,1}^{\top} & S_{\mu,1} V_{\mu} \\ 0 & V_{\mu} S_{\mu,1}^{\top} & V_{\mu} \end{pmatrix}$$

where the prior on the unconditional means is $\bar{\mu} \sim \mathrm{N}\left(\bar{m}_{\mu}, V_{\mu}\right)$. We use the Kalman filter and simulation smoothing algorithm to draw the missing values and the unconditional means jointly.

Appendix G Modeling implications

Appendix G.1 Recursive preferences

The recursive utility implies:

$$\begin{aligned}
\mathcal{M}_{t,t+1} \quad &= \quad \beta e^{(\rho-1)\Delta c_{t+1}} \left[\frac{e^{\Delta c_{t+1} + u_{t+1}}}{\mu_t\left(e^{\Delta c_{t+1} + u_{t+1}}\right)}\right]^{\alpha-\rho}, \\
\widehat{\mathcal{M}}_{t,t+1} \quad &= \quad \beta e^{(\widehat{\rho}-1)\Delta \widehat{c}_{t+1}} \left[\frac{e^{\Delta \widehat{c}_{t+1} + \widehat{u}_{t+1}}}{\mu_t\left(e^{\Delta \widehat{c}_{t+1} + \widehat{u}_{t+1}}\right)}\right]^{\alpha-\widehat{\rho}}.
\end{aligned}$$

Here $u_t = \log(U_t/C_t)$ with U_t denoting the time aggregator,

$$U_t = [(1-\beta)C_t^{\rho} + \beta\mu_t(U_{t+1})^{\rho}]^{1/\rho}.$$

The certainty equivalent is $\mu_t(x_{t+1}) = (E_t x_{t+1}^{\alpha})^{1/\alpha}$. We use the same notation for the foreign SDF, but with "hats".

The scaled log time aggregator u_t can be approximated via

$$u_t \approx b_0 + b_1 \log \mu_t \left(e^{\Delta c_{t+1} + u_{t+1}}\right). \tag{G.1}$$

To see the relation to the power utility case, note that the pricing kernel can be re-written as:

$$\begin{aligned}
\mathcal{M}_{t,t+1} \quad &= \quad \beta e^{(\alpha-1)\Delta c_{t+1}} \left[\frac{e^{u_{t+1}}}{\mu_t\left(e^{\Delta c_{t+1} + u_{t+1}}\right)}\right]^{\alpha-\rho} \\
&\approx \quad \beta' e^{(\alpha-1)\Delta c_{t+1}} e^{(\alpha-\rho)(u_{t+1}-u_t)}.
\end{aligned}$$

The approximation holds if $b_0 = 0$ and $b_1 = 1$ in equation (G.1). Hansen (2012) discusses this case and shows that, given the stationarity of u_t, the long-run properties of the SDF are the same as those of the power utility case. Thus, while a generic recursive SDF does not literally behave like the power utility one in the long run, that relationship is very close because in practice $b_0 \approx 0$ and $b_1 \approx 1$.

Here we are more explicit about the dynamics of the conditional variance of consumption growth. We assume that $\sigma_t^2 \sim ARG(c, \varphi, \kappa)$ and $\widehat{\sigma}_t^2 \sim ARG(\widehat{c}, \widehat{\varphi}, \widehat{\kappa})$. Thus,

$$k_t(s, \sigma_{t+1}^2) = \varphi s(1 - sc)^{-1}\sigma_t^2 - \kappa \log(1 - sc).$$

Occasionally, we write the variance process in the dynamic form

$$\sigma_{t+1}^2 = c\kappa + \varphi\sigma_t^2 + (c^2\kappa + 2c\varphi\sigma_t^2)^{1/2}\omega_{t+1}$$

to emphasize the iid innovations ω_t with zero mean and unit variance.

To solve for the aggregator, guess

$$\begin{aligned} u_{t+1} &= u_0 + u_c c_{t+1} + u_\sigma \sigma_{t+1}^2 + u_{\widehat{\sigma}} \widehat{\sigma}_{t+1}^2 \\ &= u_0 + u_c(\mu - \widehat{\mu}) + u_c(1 + \theta - \widehat{\theta})\Delta_c c_t + u_c \sigma_t \varepsilon_{t+1} - u_c \widehat{\sigma}_t \widehat{\varepsilon}_{t+1} + u_\sigma \sigma_{t+1}^2 + u_{\widehat{\sigma}} \widehat{\sigma}_{t+1}^2. \end{aligned}$$

Then,

$$\Delta c_{t+1} + u_{t+1} = \mu + u_0 + u_c(\mu - \widehat{\mu}) + [\theta + u_c(1 + \theta - \widehat{\theta})]\Delta_c c_t + (u_c + 1)\sigma_t \varepsilon_{t+1} - u_c \widehat{\sigma}_t \widehat{\varepsilon}_{t+1} + u_\sigma \sigma_{t+1}^2 + u_{\widehat{\sigma}} \widehat{\sigma}_{t+1}^2,$$

and

$$\begin{aligned} \log \mu_t \left(e^{\Delta c_{t+1} + u_{t+1}}\right) &= \mu + u_0 + u_c(\mu - \widehat{\mu}) + [\theta + u_c(1 + \theta - \widehat{\theta})]\Delta_c c_t + \alpha(u_c + 1)^2 \sigma_t^2/2 + \alpha u_c^2 \widehat{\sigma}_t^2/2 \\ &+ k_t(\alpha u_\sigma, \sigma_{t+1}^2)/\alpha + k_t(\alpha u_{\widehat{\sigma}}, \widehat{\sigma}_{t+1}^2)/\alpha. \end{aligned}$$

Lining up terms in u_t and using the approximation (G.1), we get

$$u_c = b_1[\theta + u_c(1 + \theta - \widehat{\theta})]$$

and

$$\begin{aligned} u_\sigma &= b_1[\alpha u_c^2/2 + \varphi u_\sigma(1 - \alpha u_\sigma c)^{-1}], \\ u_{\widehat{\sigma}} &= b_1[\alpha u_c^2/2 + \widehat{\varphi} u_{\widehat{\sigma}}(1 - \alpha \widehat{u}_\sigma \widehat{c})^{-1}]. \end{aligned}$$

Thus,

$$u_c = b_1\theta/(1 - b_1(1 + \theta - \widehat{\theta})) = \theta/(\widehat{\theta} - \theta),$$

and

$$\begin{aligned} u_\sigma &= (2c\alpha)^{-1}[1 - \varphi - \sqrt{(1 - \varphi)^2 - 2c\alpha^2 u_c^2}], \\ u_{\widehat{\sigma}} &= (2\widehat{c}\alpha)^{-1}[1 - \widehat{\varphi} - \sqrt{(1 - \widehat{\varphi})^2 - 2\widehat{c}\alpha^2 u_c^2}]. \end{aligned}$$

Given the solution for u_t, we have

$$\begin{aligned} \Delta c_{t+1} + u_{t+1} - \log \mu_t \left(e^{\Delta c_{t+1} + u_{t+1}}\right) &= (u_c + 1)\sigma_t \varepsilon_{t+1} - u_c \widehat{\sigma}_t \widehat{\varepsilon}_{t+1} + u_\sigma \sigma_{t+1}^2 + u_{\widehat{\sigma}} \widehat{\sigma}_{t+1}^2 \\ &- \alpha(u_c + 1)^2 \sigma_t^2/2 - \alpha u_c^2 \widehat{\sigma}_t^2/2 - k_t(\alpha u_\sigma, \sigma_{t+1}^2)/\alpha - k_t(\alpha u_{\widehat{\sigma}}, \widehat{\sigma}_{t+1}^2)/\alpha \end{aligned}$$

and

$$\begin{aligned}
\log \mathcal{M}_{t,t+1} &= \log\beta + (\rho-1)\mu + (\rho-1)\theta\Delta_c c_t + (\alpha-1+u_c(\alpha-\rho))\sigma_t\varepsilon_{t+1} - (\alpha-\rho)u_c\widehat{\sigma}_t\widehat{\varepsilon}_{t+1} \\
&+ (\alpha-\rho)[u_\sigma\sigma_{t+1}^2 + u_{\widehat{\sigma}}\widehat{\sigma}_{t+1}^2 - (u_c+1)^2\sigma_t^2/2 - u_c^2\widehat{\sigma}_t^2/2 - k_t(u_\sigma,\sigma_{t+1}^2) - k_t(u_{\widehat{\sigma}},\widehat{\sigma}_{t+1}^2)] \\
&= \log\beta + [(\rho-1)\mu - (\alpha-\rho)(u_\sigma c\kappa + u_{\widehat{\sigma}}\widehat{c}\widehat{\kappa} + \kappa\alpha^{-1}\log(1-u_\sigma c\alpha) + \widehat{\kappa}\alpha^{-1}\log(1-\widehat{u}_\sigma\widehat{c}\alpha))] \\
&+ (\rho-1)\theta\Delta_c c_t + (\alpha-\rho)\sigma_t^2[u_\sigma\varphi - \alpha(u_c+1)^2/2 - \varphi u_\sigma(1-u_\sigma c\alpha)^{-1}] \\
&+ (\alpha-\rho)\widehat{\sigma}_t^2[u_{\widehat{\sigma}}\widehat{\varphi} - \alpha u_c^2/2 - \widehat{\varphi}\widehat{u}_\sigma(1-\widehat{u}_\sigma\widehat{c}\alpha)^{-1}] \\
&+ (\alpha-1+u_c(\alpha-\rho))\ \sigma_t\varepsilon_{t+1} - (\alpha-\rho)u_c\widehat{\sigma}_t\widehat{\varepsilon}_{t+1} \\
&+ (\alpha-\rho)[u_\sigma(c^2\kappa + 2c\varphi\sigma_t^2)^{1/2}\omega_{t+1} + u_{\widehat{\sigma}}(\widehat{c}^2\kappa + 2\widehat{c}\widehat{\varphi}\widehat{\sigma}_t^2)^{1/2}\widehat{\omega}_{t+1}].
\end{aligned}$$

The spot interest rate, $r_t = -\log E_t\mathcal{M}_{t,t+1}$, depends on three state variables: $\Delta_c c_t$, σ_t^2, and $\widehat{\sigma_t}^2$. That is the case for the foreign interest rate, $\widehat{r}_t$, as well. As a result, the interest rate differential $\Delta_c r_t$ does not summarize the entire state vector. Thus, qualitatively, the model is capable of producing extra predictors of excess currency returns.

Next, we consider implications for the conditional variance of the pricing kernel:

$$\begin{aligned}
var_t(\log\mathcal{M}_{t,t+1}) &= (\alpha-\rho)^2\kappa(u_\sigma^2c^2 + u_{\widehat{\sigma}}^2\widehat{c}^2) \\
&+ [(\alpha-1+u_c(\alpha-\rho))^2 + 2(\alpha-\rho)^2u_\sigma^2c\varphi]\sigma_t^2 \\
&+ [(\alpha-\rho)^2u_c^2 + 2(\alpha-\rho)^2u_{\widehat{\sigma}}^2\widehat{c}\widehat{\varphi}]\widehat{\sigma}_t^2.
\end{aligned}$$

Thus, conditional variation of the SDF is controlled by two out of the three state variables. That is a generic feature of the models with recursive preferences: the conditional volatility of the SDF is affected only by those variables driving the conditional volatility of the innovations to the consumption process. Thus, variation in the pricing kernel, a.k.a. the maximal risk premium is driven by variables that are associated with macroeconomic activity at the business cycle frequency (Segal, Shaliastovich, and Yaron, 2015). We cannot replicate our empirical result that the RER appears in $var_t(\log\mathcal{M}_{t,t+1})$.

Appendix G.2 Habits

The habit-based utility implies:

$$\begin{aligned}
\mathcal{M}_{t,t+1} &= \beta e^{(\alpha-1)\Delta c_{t+1}}\left[\frac{H_{t+1}}{H_t}\right]^{\alpha-1}, \\
\widehat{\mathcal{M}}_{t,t+1} &= \beta e^{(\alpha-1)\Delta\widehat{c}_{t+1}}\left[\frac{\widehat{H}_{t+1}}{\widehat{H}_t}\right]^{\alpha-1}.
\end{aligned}$$

Here H_t is the surplus consumption ratio. We use the same notation for the foreign SDF, but with "hats". As Bansal and Lehmann (1997) point out on the basis of this representation, if H_t is stationary the long-run properties of the habit-based SDF are the same as those of power utility.

We explore the version of H_t proposed in Campbell and Cochrane (1999) and applied in the international context by Verdelhan (2010). In these models the conditional volatility of consumption growth is assumed to be constant, $\sigma_t = \sigma$, $\widehat{\sigma}_t = \widehat{\sigma}$, because the time-variation in risk premiums emanates from varying volatility of log H_t :

$$h_{t+1} = (1-\phi)h + \phi h_t + v(h_t)\sigma\varepsilon_{t+1}.$$

Following Verdelhan (2010), we assume that the foreign h_t is controlled by the same parameters, but different shocks:

$$\widehat{h}_{t+1} = (1-\phi)h + \phi\widehat{h}_t + v(\widehat{h}_t)\widehat{\sigma}\widehat{\varepsilon}_{t+1}.$$

The RER is

$$\mathcal{E}_t = \mathcal{E}_0 e^{(\alpha-1)(c_0+h_0-\widehat{h}_0)} \cdot e^{(1-\alpha)(\Delta_c c_t + h_t - \widehat{h}_t)}.$$

The domestic and foreign interest rates are, respectively,

$$\begin{aligned} r_t &= -\log\beta + (1-\alpha)E_t\Delta c_{t+1} + (1-\alpha)E_t\Delta h_{t+1}, \\ \hat{r}_t &= -\log\beta + (1-\alpha)E_t\Delta \hat{c}_{t+1} + (1-\alpha)E_t\Delta \widehat{h}_{t+1}. \end{aligned}$$

Thus, the interest rate differential

$$\begin{aligned} \Delta_c r_t &= (1-\alpha)E_t\Delta c_{t+1} + (1-\alpha)E_t(\Delta h_{t+1} - \Delta\widehat{h}_{t+1}) \\ &= (1-\alpha)[\mu - \widehat{\mu} + (\theta - \widehat{\theta})\Delta_c c_t] + (1-\alpha)(1-\phi)(\widehat{h}_t - h_t). \end{aligned}$$

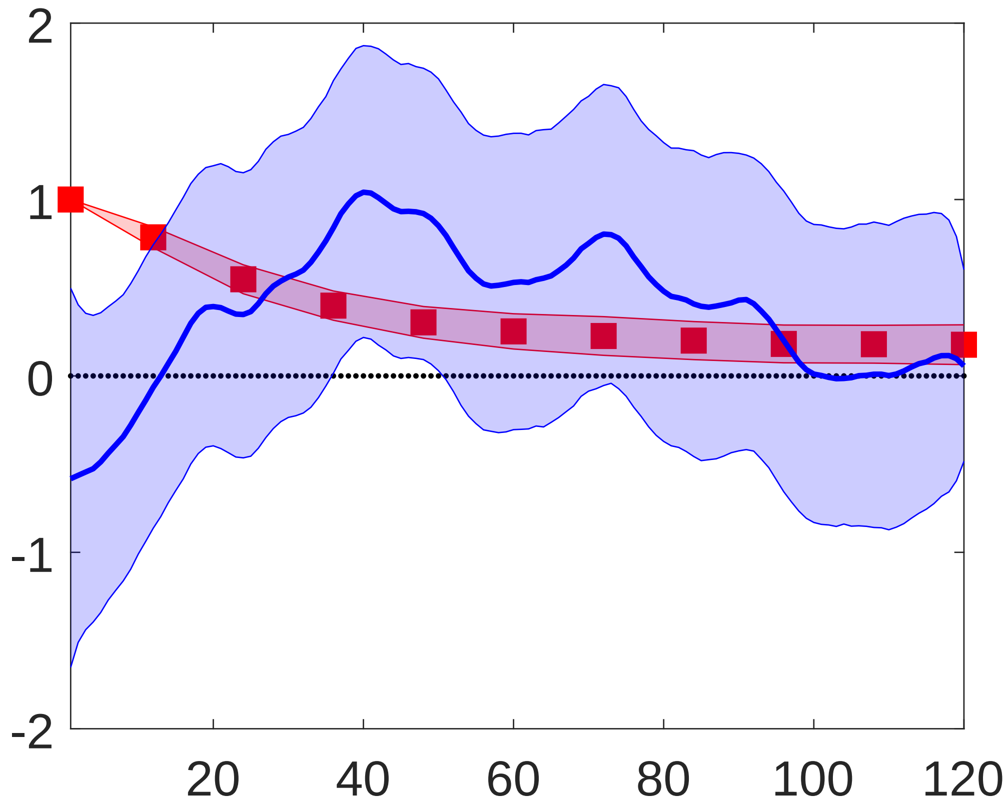
Here the RER and the IRD summarize $\Delta_c c_t$ and $h_t - \widehat{h}_t$ so there is a possibility that the RER could serve as an extra forecasting variable for excess currency returns.

The conditional variance of the log SDF

$$var_t(\log\mathcal{M}_{t,t+1}) = (\alpha-1)^2\sigma^2(1+v(h_t))^2$$

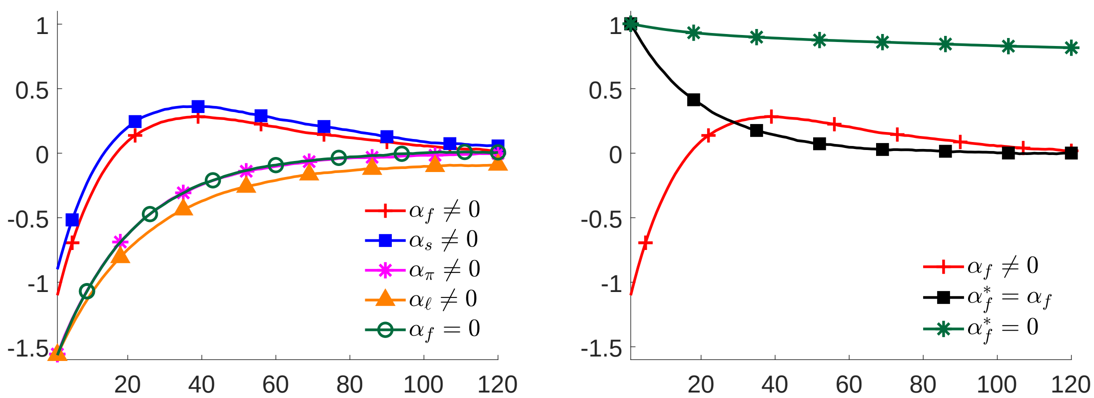
is related to h_t, but not to $\Delta_c c_t$. That is similar to the case of recursive preferences. We cannot replicate our empirical result that the RER appears in $var_t(\log\mathcal{M}_{t,t+1})$.

Figure 1
True and risk-adjusted forecasts of depreciation rates



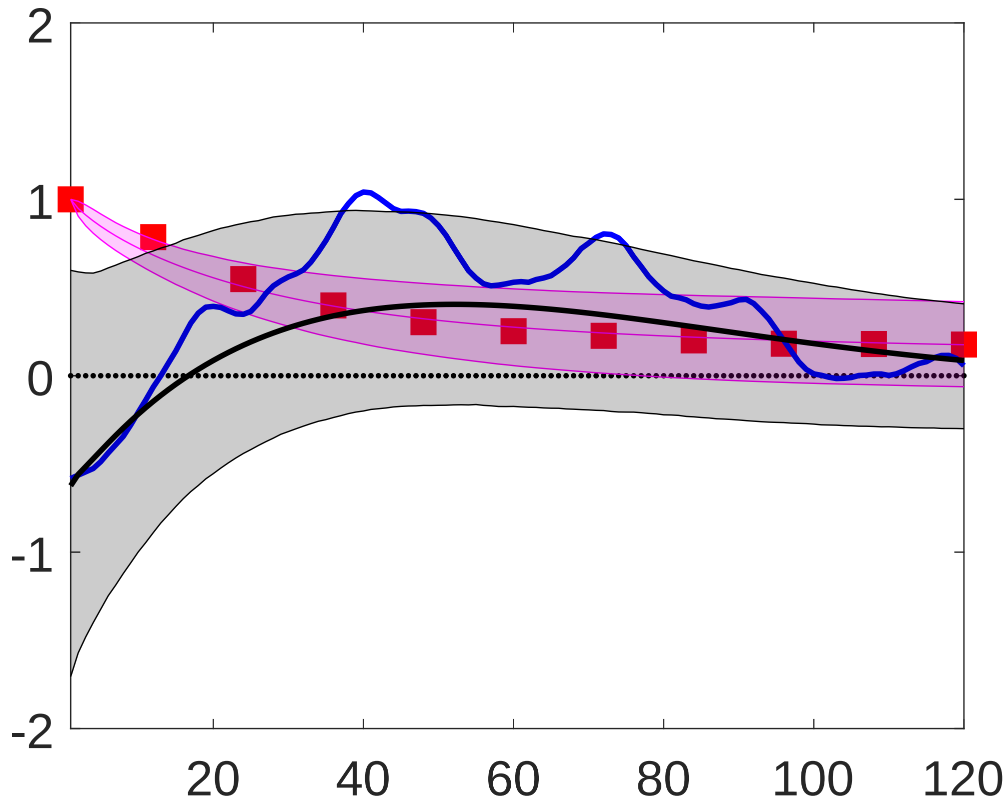
Notes. We use data on spot and forward FX rates between the U.S. dollar and currencies of the U.K., Japan, Canada, and Eurozone to document their relationship to the current IRD. We report regression slopes averaged across countries for the period from January 1983 to December 2015. For the risk-adjusted forecasts (red dots), the dependent variable corresponds to expected average annual change in FX rates. For the real-world forecasts (blue line), the dependent variable corresponds to monthly changes in FX rates.

Figure 2
True and risk-adjusted forecasts of depreciation rates in a simple U.K. model



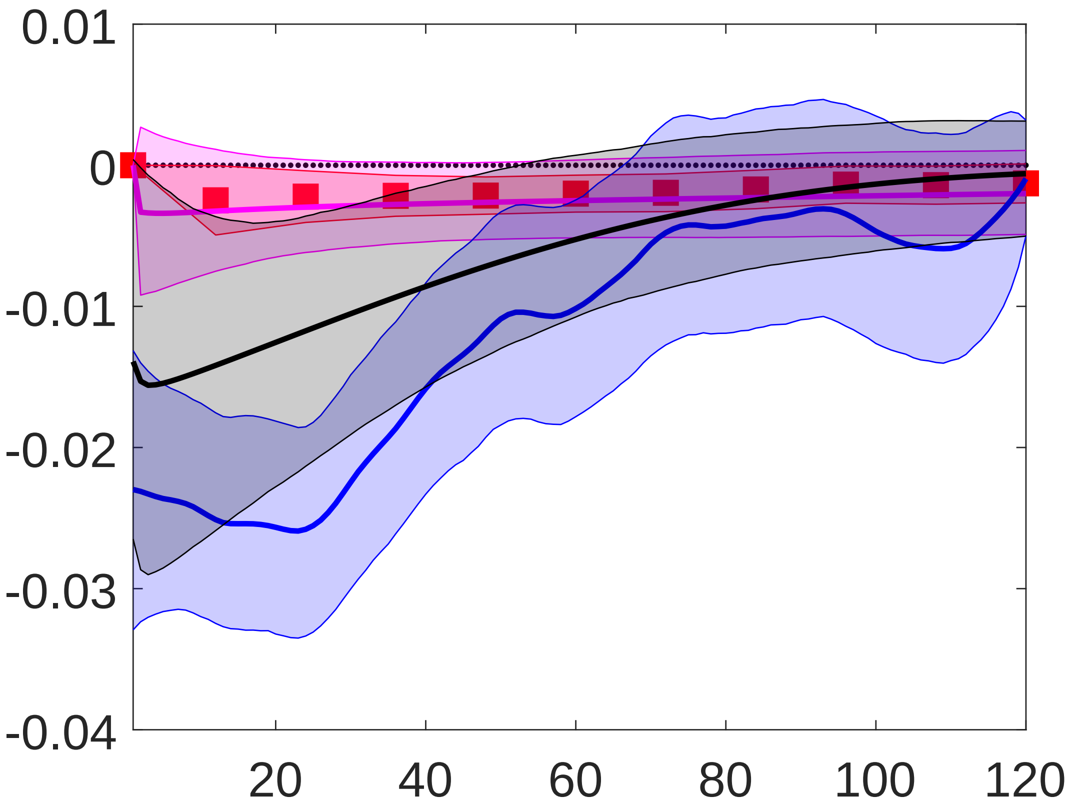
Notes. We explore variations in the multihorizon pattern in the UIP regressions implied by a simple bilateral VECM model of U.S. and U.K. On the left panel variations are associated with variations in the values of elements of α_f. VAR corresponds to all $\alpha_f = 0$. On the right panel, we consider various scenarios of the values of α_f^*. The red line with crosses, $\alpha_f \neq 0$, is the same across the two panels and represents the implications of the estimated VECM.

Figure 3
True and risk-adjusted cross-covariances of depreciation rates and the interest rate differential



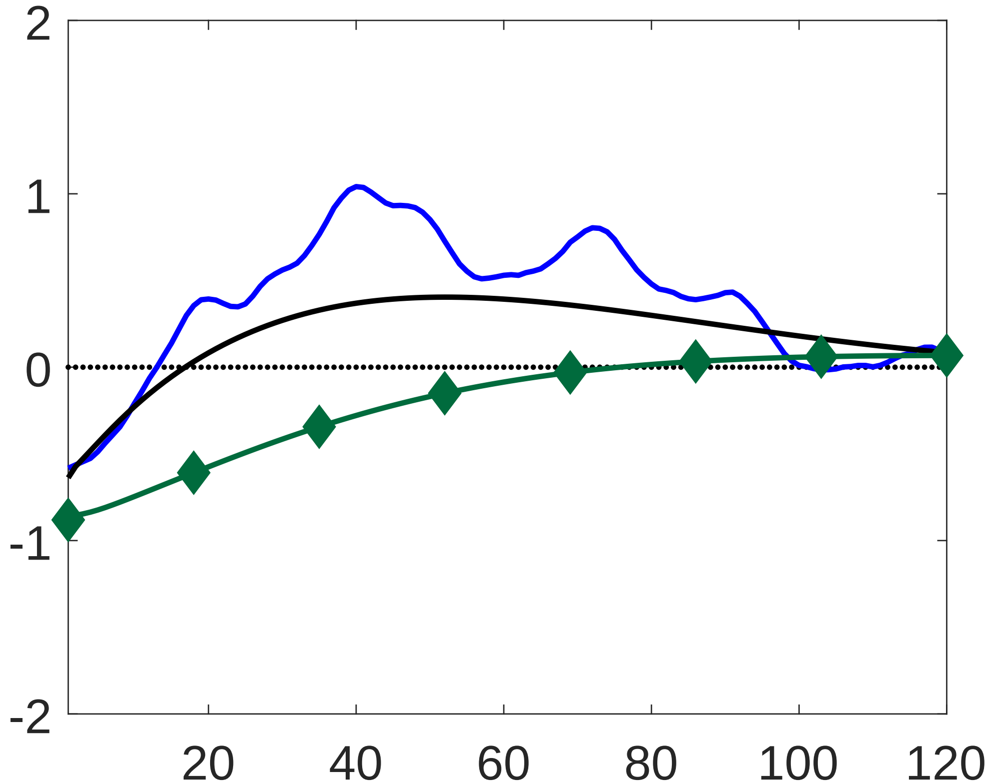
Notes. We report multi-horizon UIP regression patterns in the data (blue line) and in the model (black line) as cross-country average regression slopes. For the risk-adjusted forecasts (red dots), the dependent variable corresponds to expected average annual change in FX rates. For the real-world forecasts, the dependent variable corresponds to monthly changes in FX rates.

Figure 4
True and risk-adjusted cross-covariances of depreciation rates and the real exchange rate



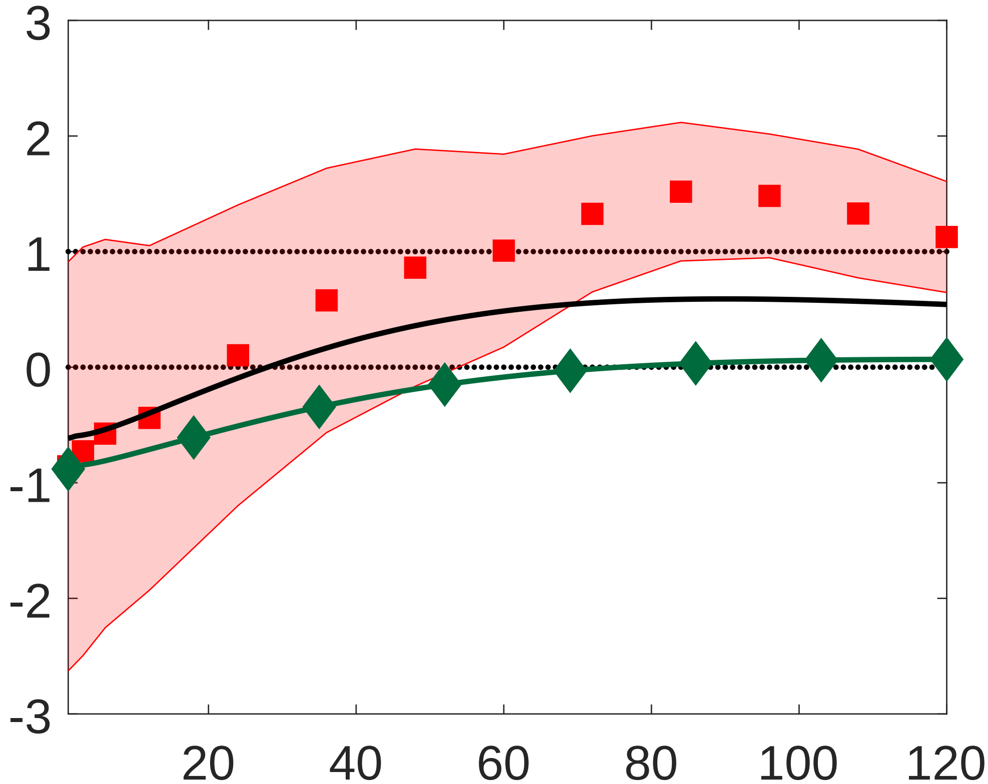
Notes. We report multi-horizon patterns of regressing future nominal depreciation rates [risk-adjusted forecasts of depreciation rates] on the current real exchange rate in the data (blue line [magenta line]) and in the model (black line [red dots]) as cross-country average regression slopes. For the risk-adjusted forecasts, the dependent variable corresponds to expected average annual change in FX rates. For the real-world forecasts, the dependent variable corresponds to monthly changes in FX rates.

Figure 5
True and risk-adjusted forecasts of depreciation rates in a VAR model



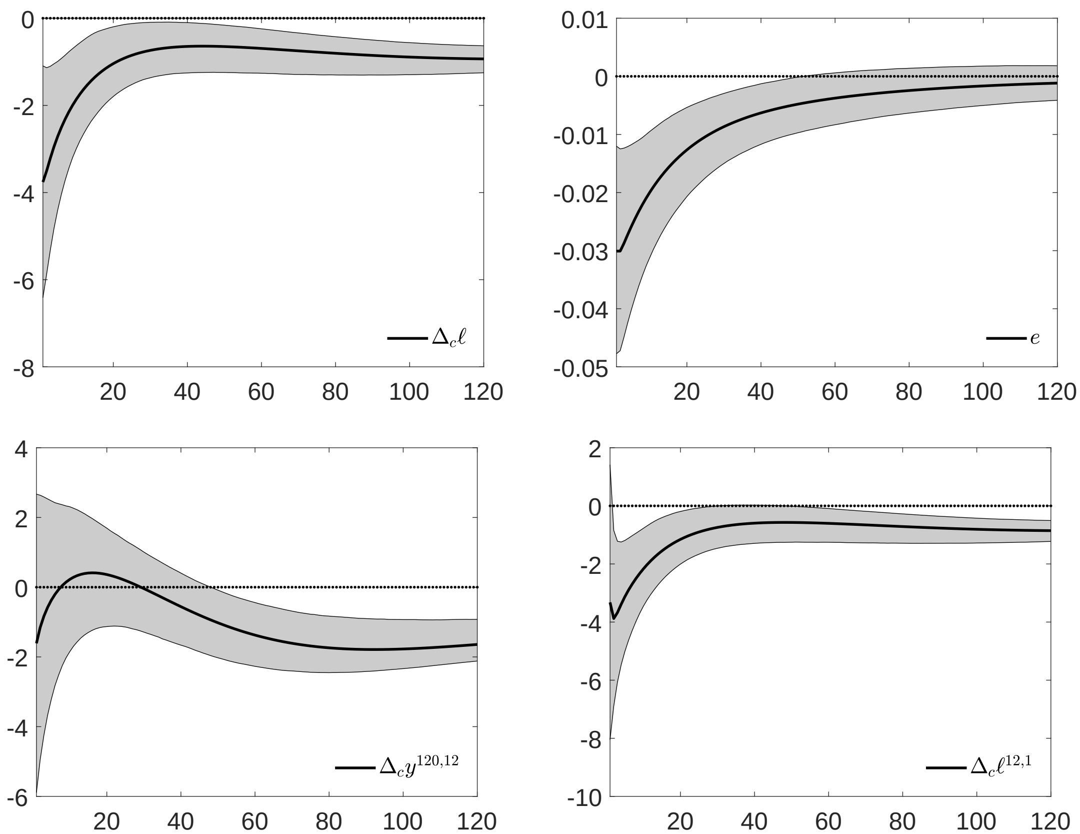
Notes. We compare the multihorizon pattern in the UIP regressions in the VAR model (green line) to those from the VECM model (black line) and in the data.

Figure 6
Long-run UIP in the VECM and VAR models



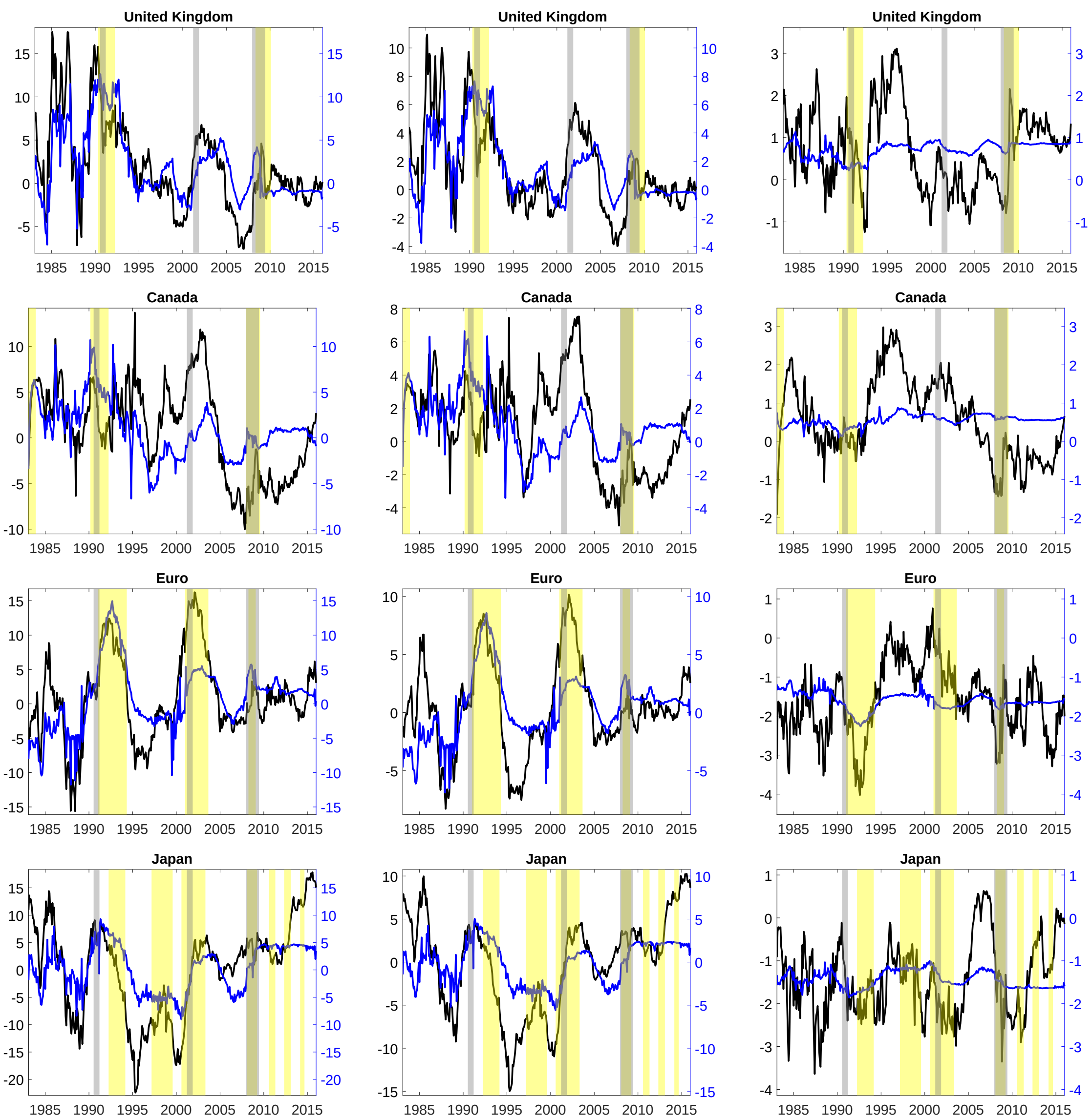
Notes. We compare the multihorizon pattern in the long-run UIP regressions in the VAR model (green line) to those from the VECM model (black line) and in the data (red squares).

Figure 7
Term structure of factor loadings in the nominal risk premiums



Notes. We report multi-horizon currency risk premium factor loadings as cross-country averages of $\kappa_{n,x}$ from $rps_t^n = \kappa_{n,0} + \kappa_{n,x}^\top x_t$. We omit the results for factors whose loadings are insignificant across all horizons.

Figure 8
Term structure of nominal and UIP-based risk premiums



Notes. The black line shows the nominal currency risk premium, rps_t^n, for horizons $n = 1, 12$, and 60 months. The blue line shows the projection of this premium on the IRD, $\Delta_c \ell_t$. Vertical gray bars are U.S. recessions while vertical yellow bars are foreign recessions in each respective country.

Table 1: Posterior mean and stand. dev. of a two country model with the U.S.-U.K.

x_t		Δs_t	$\Delta_c \pi_t$	ℓ_t	$y_t^{120,12}$	c_t	$\Delta_c \ell_t$	$\Delta_c y_t^{120,12}$	$\Delta_c \ell_t^{12,1}$	e_t
	$\bar{\mu}_x \times 1200$		Φ_x							
Δs_t	-0.224	0.065	0.060	-0.792	-2.968	-2.187	-3.430	-0.179	-3.755	-0.053
	(0.558)	(0.049)	(0.184)	(0.908)	(2.059)	(3.166)	(1.571)	(2.319)	(3.049)	(0.015)
$\Delta_c \pi_t$	-0.224	0.019	0.210	-0.117	-0.067	0.899	1.095	0.510	1.959	0.008
	(0.558)	(0.008)	(0.046)	(0.163)	(0.368)	(0.584)	(0.295)	(0.416)	(0.584)	(0.003)
ℓ_t	4.407	6.06e-04	0.006	0.991	-0.057	-0.153	0.023	-0.022	0.103	1.44e-04
	(0.663)	(7.64e-04)	(0.005)	(0.015)	(0.033)	(0.053)	(0.026)	(0.037)	(0.052)	(2.62e-04)
$y_t^{120,12}$	1.686	4.56e-05	-9.87e-05	-0.017	0.928	-0.009	-0.025	0.004	-0.020	-3.42e-04
	(0.271)	(3.50e-04)	(0.002)	(0.007)	(0.016)	(0.028)	(0.013)	(0.018)	(0.027)	(1.25e-04)
c_t	0.648	8.12e-05	-0.005	0.065	0.027	0.537	0.028	0.027	0.050	4.75e-04
	(0.099)	(5.46e-04)	(0.003)	(0.011)	(0.025)	(0.041)	(0.020)	(0.028)	(0.039)	(1.93e-04)
$\Delta_c \ell_t$	-1.726	0.001	-7.72e-04	-0.031	0.019	-0.012	1.032	-0.132	0.582	0.001
	(0.396)	(8.99e-04)	(0.006)	(0.016)	(0.037)	(0.056)	(0.028)	(0.041)	(0.055)	(3.01e-04)
$\Delta_c y_t^{120,12}$	0.983	3.12e-04	0.003	0.002	-0.027	0.010	-0.009	0.972	-0.037	-2.61e-04
	(0.255)	(4.18e-04)	(0.003)	(0.008)	(0.019)	(0.031)	(0.015)	(0.021)	(0.030)	(1.47e-04)
$\Delta_c \ell_t^{12,1}$	0.353	-0.001	0.003	0.027	0.011	-0.018	-0.070	0.072	0.468	-6.71e-04
	(0.107)	(7.08e-04)	(0.004)	(0.013)	(0.030)	(0.046)	(0.023)	(0.033)	(0.045)	(2.39e-04)
e_t	0.462†	0.045	-0.150	-0.674	-2.900	-3.086	-4.524	-0.689	-5.714	0.939
	(0.031)	(0.050)	(0.188)	(0.914)	(2.072)	(3.196)	(1.584)	(2.333)	(3.076)	(0.015)
x_t	$\delta_{\widehat{c},x}$		$\Sigma_x \times \sqrt{12} \times 100$							
Δs_t	0.003	10.115	0	0	0	0	0	0	0	
	(0.001)	(0.370)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	
$\Delta_c \pi_t$	-0.010	0.111	1.674	0	0	0	0	0	0	
	(0.004)	(0.086)	(0.064)	(—)	(—)	(—)	(—)	(—)	(—)	
ℓ_t	0.071	-0.014	0.001	0.151	0	0	0	0	0	
	(0.017)	(0.007)	(0.007)	(0.005)	(—)	(—)	(—)	(—)	(—)	
$y_t^{120,12}$	0.084	-0.001	0.004	6.28e-04	0.069	0	0	0	0	
	(0.034)	(0.003)	(0.004)	(0.003)	(0.002)	(—)	(—)	(—)	(—)	
c_t	-0.504	0.004	-0.018	0.041	0.007	0.098	0	0	0	
	(0.070)	(0.006)	(0.006)	(0.005)	(0.005)	(0.003)	(—)	(—)	(—)	
c_t	0.083	0.007	-0.007	0.012	-0.018	0.020	0.172	0	0	
	(0.026)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.006)	(—)	(—)	
$\Delta_c y_t^{120,12}$	-0.119	0.005	0.011	-2.04e-04	0.032	0.001	-0.031	0.068	0	
	(0.035)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(—)	
$\Delta_c \ell_t^{12,1}$	0.181	-0.018	-0.005	-0.014	-0.001	-0.009	-0.097	-0.047	0.080	
	(0.054)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.006)	(0.004)	(0.003)	
e_t	5.51e-04	10.004	-1.674	0	0	0	0	0	0	
	(1.87e-04)	(0.376)	(0.064)	(—)	(—)	(—)	(—)	(—)	(—)	
x_t	$\bar{\mu}_x^* \times 1200$		Φ_x^*							
Δs_t	-3.546	0	0	0	0	0	1	0	0	0
	(2.771)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
$\Delta_c \pi_t$	-3.546	0.035	0.587	0.336	-0.807	0.632	3.024	-0.714	4.503	0.016
	(2.771)	(0.023)	(0.036)	(0.211)	(0.487)	(0.558)	(0.392)	(0.580)	(0.599)	(0.002)
ℓ_t	20.235	5.75e-04	-0.009	1.023	0.059	-0.177	0.089	-0.035	0.176	3.40e-04
	(1.313)	(4.40e-04)	(0.001)	(0.006)	(0.014)	(0.026)	(0.012)	(0.015)	(0.025)	(7.74e-05)
$y_t^{120,12}$	-0.620	3.35e-05	0.001	-0.004	0.972	-0.047	-0.009	7.32e-04	-0.018	-8.40e-05
	(0.040)	(4.24e-05)	(1.60e-04)	(0.001)	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(9.89e-06)
c_t	0.525	8.18e-04	-0.008	0.032	0.054	0.460	0.094	-0.040	0.194	1.04e-04
	(0.391)	(5.98e-04)	(0.002)	(0.007)	(0.017)	(0.030)	(0.017)	(0.017)	(0.030)	(1.02e-04)
$\Delta_c \ell_t$	-3.034	0.001	-0.002	0.005	-0.003	-0.033	1.023	-0.078	0.489	1.16e-04
	(2.770)	(4.39e-04)	(9.03e-04)	(0.002)	(0.005)	(0.016)	(0.006)	(0.007)	(0.018)	(3.16e-05)
$\Delta_c y_t^{120,12}$	0.249	2.73e-04	-8.13e-04	0.007	0.011	-0.117	0.018	0.992	-0.072	1.63e-05
	(0.067)	(1.09e-04)	(2.99e-04)	(0.001)	(0.003)	(0.005)	(0.002)	(0.004)	(0.005)	(1.74e-05)
$\Delta_c \ell_t^{12,1}$	0.009	-0.001	0.002	-0.005	0.004	0.031	-0.025	0.097	0.613	-1.31e-04
	(0.006)	(4.49e-04)	(9.60e-04)	(0.002)	(0.006)	(0.017)	(0.007)	(0.007)	(0.019)	(3.44e-05)
e_t	0.126†	-0.035	-0.587	-0.336	0.807	-0.632	-2.024	0.714	-4.503	0.984
	(0.375)	(0.023)	(0.036)	(0.211)	(0.487)	(0.558)	(0.392)	(0.580)	(0.599)	(0.002)

Posterior mean and standard deviation for a two country model of the U.S. and United Kingdom. We report unconditional means $\bar{\mu}_x^*, \bar{\mu}_x$, autocovariances matrices Φ_x^*, Φ_x, scale matrix Σ_x and the vector of loadings $\delta_{\widehat{c},x}$. The symbol † indicates that the unconditional means of the real exchange rates are not multiplied by 1200 (annualization and conversion to percent is not applicable).

Table 2: Posterior mean and stand. dev. of a two country model with the U.S.-Canada

x_t		Δs_t	$\Delta_c \pi_t$	ℓ_t	$y_t^{120,12}$	c_t	$\Delta_c \ell_t$	$\Delta_c y_t^{120,12}$	$\Delta_c \ell_t^{12,1}$	e_t
	$\bar{\mu}_x \times 1200$		Φ_x							
Δs_t	0.212	-0.061	-0.006	-0.529	-1.426	0.345	-2.103	-0.676	-2.456	-0.016
	(0.266)	(0.050)	(0.187)	(0.733)	(1.553)	(2.620)	(1.622)	(2.362)	(2.667)	(0.008)
$\Delta_c \pi_t$	0.212	0.029	0.001	-0.142	-0.412	-0.445	-0.671	-1.281	-0.788	0.002
	(0.266)	(0.008)	(0.043)	(0.119)	(0.246)	(0.413)	(0.265)	(0.389)	(0.441)	(0.001)
ℓ_t	4.452	2.76e-04	0.002	0.975	-0.081	-0.107	-0.042	-0.056	-0.022	8.33e-05
	(0.699)	(0.001)	(0.007)	(0.015)	(0.032)	(0.052)	(0.034)	(0.048)	(0.056)	(1.92e-04)
$y_t^{120,12}$	1.656	9.50e-04	0.004	-0.020	0.936	0.009	-0.042	-0.036	0.017	-3.55e-05
	(0.264)	(4.92e-04)	(0.003)	(0.008)	(0.016)	(0.029)	(0.018)	(0.027)	(0.030)	(9.04e-05)
c_t	0.654	2.89e-04	9.65e-04	0.074	0.021	0.525	0.061	0.092	0.038	1.46e-04
	(0.101)	(7.69e-04)	(0.005)	(0.012)	(0.025)	(0.042)	(0.028)	(0.040)	(0.045)	(1.45e-04)
$\Delta_c \ell_t$	-0.825	8.57e-04	-0.001	-0.053	-0.034	0.040	0.881	-0.266	0.366	-1.22e-04
	(0.369)	(0.001)	(0.007)	(0.017)	(0.034)	(0.054)	(0.033)	(0.050)	(0.053)	(1.97e-04)
$\Delta_c y_t^{120,12}$	0.496	3.34e-04	0.001	-0.008	-0.021	0.013	0.001	0.949	5.23e-04	1.40e-04
	(0.204)	(5.82e-04)	(0.004)	(0.010)	(0.019)	(0.033)	(0.021)	(0.031)	(0.033)	(1.06e-04)
$\Delta_c \ell_t^{12,1}$	0.065	-0.002	0.003	0.027	0.014	-0.017	-0.007	0.148	0.488	3.33e-05
	(0.077)	(8.93e-04)	(0.006)	(0.015)	(0.028)	(0.047)	(0.029)	(0.043)	(0.046)	(1.66e-04)
e_t	$-0.171^\dagger$	-0.090	-0.008	-0.387	-1.014	0.790	-1.431	0.605	-1.669	0.982
	(0.043)	(0.051)	(0.192)	(0.741)	(1.569)	(2.650)	(1.644)	(2.392)	(2.703)	(0.008)
x_t	$\delta_{\widehat{c},x}$		$\Sigma_x \times \sqrt{12} \times 100$							
Δs_t	3.15e-04	7.249	0	0	0	0	0	0	0	
	(0.004)	(0.262)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	
$\Delta_c \pi_t$	-0.003	0.009	1.062	0	0	0	0	0	0	
	(0.014)	(0.055)	(0.039)	(—)	(—)	(—)	(—)	(—)	(—)	
ℓ_t	0.020	-0.002	-0.003	0.148	0	0	0	0	0	
	(0.049)	(0.007)	(0.007)	(0.004)	(—)	(—)	(—)	(—)	(—)	
$y_t^{120,12}$	-0.008	-0.001	0.010	-0.002	0.068	0	0	0	0	
	(0.037)	(0.003)	(0.003)	(0.003)	(0.002)	(—)	(—)	(—)	(—)	
c_t	0.248	-0.008	-0.003	0.039	0.006	0.100	0	0	0	
	(0.155)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(—)	(—)	(—)	
$\Delta_c \ell_t$	-0.070	0.016	0.007	0.030	-0.002	0.025	0.141	0	0	
	(0.029)	(0.008)	(0.008)	(0.013)	(0.008)	(0.008)	(0.005)	(—)	(—)	
$\Delta_c y_t^{120,12}$	-0.016	-0.006	3.02e-04	0.021	0.034	0.002	-0.038	0.058	0	
	(0.050)	(0.004)	(0.004)	(0.007)	(0.004)	(0.004)	(0.003)	(0.002)	(—)	
$\Delta_c \ell_t^{12,1}$	-0.281	-0.010	-0.008	-0.039	-0.010	-0.021	-0.072	-0.047	0.074	
	(0.072)	(0.006)	(0.007)	(0.011)	(0.007)	(0.007)	(0.005)	(0.004)	(0.003)	
e_t	1.99e-04	7.240	-1.062	0	0	0	0	0	0	
	(9.72e-05)	(0.267)	(0.039)	(—)	(—)	(—)	(—)	(—)	(—)	
x_t	$\bar{\mu}_x^* \times 1200$		Φ_x^*							
Δs_t	24.581	0	0	0	0	0	1	0	0	0
	(3.139)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
$\Delta_c \pi_t$	24.581	0.067	0.158	1.913	-2.807	0.557	-1.457	-2.819	-1.241	0.006
	(3.139)	(0.033)	(0.039)	(0.351)	(0.640)	(0.255)	(0.285)	(0.656)	(0.601)	(8.01e-04)
ℓ_t	22.181	-2.13e-04	-0.006	1.018	0.040	-0.008	-0.008	-0.019	-0.008	4.03e-05
	(1.232)	(6.88e-04)	(0.003)	(0.008)	(0.012)	(0.006)	(0.005)	(0.013)	(0.010)	(2.38e-05)
$y_t^{120,12}$	-0.524	4.51e-05	4.59e-04	7.41e-04	0.977	-0.073	0.011	0.018	0.009	-7.04e-06
	(0.030)	(5.73e-05)	(2.78e-04)	(8.82e-04)	(0.002)	(7.27e-04)	(0.002)	(0.003)	(0.002)	(3.36e-06)
c_t	4.898	2.33e-04	-0.020	0.095	0.039	0.395	0.055	0.062	0.050	1.12e-04
	(0.890)	(8.70e-04)	(0.003)	(0.013)	(0.021)	(0.026)	(0.016)	(0.030)	(0.030)	(3.20e-05)
$\Delta_c \ell_t$	24.844	8.20e-04	1.00e-03	-0.008	0.010	0.032	0.999	-0.094	0.461	-2.41e-05
	(3.138)	(6.21e-04)	(0.001)	(0.005)	(0.005)	(0.019)	(0.002)	(0.008)	(0.019)	(1.16e-05)
$\Delta_c y_t^{120,12}$	0.182	5.47e-05	-1.68e-04	0.005	0.010	-0.057	0.008	1.005	-0.106	-5.45e-06
	(0.065)	(2.87e-04)	(0.001)	(0.004)	(0.003)	(0.012)	(0.002)	(0.004)	(0.005)	(7.57e-06)
$\Delta_c \ell_t^{12,1}$	0.009	-8.37e-04	-9.71e-04	0.009	-0.012	-0.035	6.11e-04	0.117	0.642	2.79e-05
	(0.004)	(6.37e-04)	(0.001)	(0.005)	(0.005)	(0.020)	(0.002)	(0.009)	(0.020)	(1.22e-05)
e_t	$-0.168^\dagger$	-0.067	-0.158	-1.913	2.807	-0.557	2.457	2.819	1.241	0.994
	(0.435)	(0.033)	(0.039)	(0.351)	(0.640)	(0.255)	(0.285)	(0.656)	(0.601)	(8.01e-04)

Posterior mean and standard deviation for a two country model of the U.S. and Canada. We report unconditional means $\bar{\mu}_x^*, \bar{\mu}_x$, autocovariances matrices Φ_x^*, Φ_x, scale matrix Σ_x and the vector of loadings $\delta_{\widehat{c},x}$. The symbol † indicates that the unconditional means of the real exchange rates are not multiplied by 1200 (annualization and conversion to percent is not applicable).

Table 3: Posterior mean and stand. dev. of a two country model with the U.S.-Euro

x_t		Δs_t	$\Delta_c \pi_t$	ℓ_t	$y_t^{120,12}$	c_t	$\Delta_c \ell_t$	$\Delta_c y_t^{120,12}$	$\Delta_c \ell_t^{12,1}$	e_t
	$\bar{\mu}_x \times 1200$		Φ_x							
Δs_t	0.917	0.133	0.097	0.039	1.640	3.423	-4.143	-3.786	-8.059	-0.037
	(0.330)	(0.049)	(0.191)	(0.840)	(2.194)	(3.151)	(1.583)	(2.537)	(2.844)	(0.011)
$\Delta_c \pi_t$	0.917	0.022	0.103	0.248	0.128	-1.303	0.585	0.473	-0.186	0.003
	(0.330)	(0.007)	(0.044)	(0.132)	(0.365)	(0.539)	(0.261)	(0.435)	(0.455)	(0.002)
ℓ_t	4.391	1.74e-04	0.006	0.981	-0.059	-0.138	0.035	-0.001	0.109	4.94e-05
	(0.624)	(7.69e-04)	(0.005)	(0.013)	(0.035)	(0.052)	(0.026)	(0.042)	(0.046)	(1.79e-04)
$y_t^{120,12}$	1.651	-3.33e-05	-6.93e-04	5.56e-05	0.974	0.020	-0.084	-0.087	-0.121	-4.38e-04
	(0.211)	(3.48e-04)	(0.003)	(0.007)	(0.018)	(0.027)	(0.014)	(0.022)	(0.024)	(8.62e-05)
c_t	0.651	4.01e-04	0.004	0.054	0.010	0.536	0.048	0.054	0.017	3.26e-04
	(0.094)	(5.53e-04)	(0.004)	(0.010)	(0.027)	(0.041)	(0.021)	(0.033)	(0.036)	(1.35e-04)
$\Delta_c \ell_t$	0.651	0.001	0.014	-0.011	0.034	0.039	0.959	-0.251	0.698	-1.05e-04
	(0.435)	(0.001)	(0.007)	(0.017)	(0.043)	(0.060)	(0.030)	(0.048)	(0.055)	(2.35e-04)
$\Delta_c y_t^{120,12}$	0.397	-8.58e-05	0.004	0.007	-0.011	-0.011	-0.056	0.927	-0.097	-2.77e-04
	(0.272)	(4.35e-04)	(0.003)	(0.008)	(0.022)	(0.033)	(0.017)	(0.027)	(0.029)	(1.06e-04)
$\Delta_c \ell_t^{12,1}$	0.149	-0.002	-0.009	-0.003	0.027	0.020	-0.012	0.117	0.375	-9.71e-06
	(0.080)	(8.59e-04)	(0.006)	(0.015)	(0.037)	(0.052)	(0.026)	(0.042)	(0.047)	(1.96e-04)
e_t	$0.221^{\dagger}$	0.111	-0.006	-0.209	1.513	4.726	-4.728	-4.259	-7.872	0.960
	(0.042)	(0.049)	(0.195)	(0.848)	(2.217)	(3.187)	(1.600)	(2.565)	(2.869)	(0.011)
x_t	$\delta_{\widehat{c},x}$		$\Sigma_x \times \sqrt{12} \times 100$							
Δs_t	-5.63e-04	9.777	0	0	0	0	0	0	0	
	(0.003)	(0.352)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	
$\Delta_c \pi_t$	-0.019	0.031	1.347	0	0	0	0	0	0	
	(0.003)	(0.069)	(0.049)	(—)	(—)	(—)	(—)	(—)	(—)	
ℓ_t	0.113	0.003	-8.41e-04	0.149	0	0	0	0	0	
	(0.016)	(0.007)	(0.007)	(0.005)	(—)	(—)	(—)	(—)	(—)	
$y_t^{120,12}$	0.158	-0.002	-0.001	0.005	0.067	0	0	0	0	
	(0.032)	(0.003)	(0.003)	(0.003)	(0.002)	(—)	(—)	(—)	(—)	
c_t	0.261	0.008	-0.004	0.038	0.008	0.099	0	0	0	
	(0.105)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(—)	(—)	(—)	
$\Delta_c \ell_t$	-0.185	4.36e-04	-0.005	0.092	0.002	0.004	0.176	0	0	
	(0.016)	(0.010)	(0.010)	(0.013)	(0.009)	(0.009)	(0.007)	(—)	(—)	
$\Delta_c y_t^{120,12}$	-0.291	-0.007	-4.11e-04	-0.013	0.060	9.02e-04	-0.004	0.056	0	
	(0.049)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(—)	
$\Delta_c \ell_t^{12,1}$	0.323	-0.012	0.008	-0.023	-0.006	-0.010	-0.144	-0.015	0.068	
	(0.114)	(0.008)	(0.008)	(0.011)	(0.008)	(0.008)	(0.007)	(0.004)	(0.002)	
e_t	-1.49e-04	9.746	-1.347	0	0	0	0	0	0	
	(3.90e-05)	(0.356)	(0.049)	(—)	(—)	(—)	(—)	(—)	(—)	
x_t	$\bar{\mu}_x^* \times 1200$		Φ_x^*							
Δs_t	9.658	0	0	0	0	0	1	0	0	0
	(1.394)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
$\Delta_c \pi_t$	9.658	-0.068	0.315	3.581	0.095	-0.308	1.144	1.757	-1.271	0.004
	(1.394)	(0.028)	(0.038)	(0.453)	(0.936)	(0.811)	(0.207)	(0.683)	(0.631)	(3.77e-04)
ℓ_t	20.695	-2.14e-04	0.007	0.968	0.064	-0.041	-0.017	-0.062	0.125	-2.69e-05
	(0.989)	(4.51e-04)	(0.001)	(0.009)	(0.011)	(0.025)	(0.003)	(0.012)	(0.025)	(1.11e-05)
$y_t^{120,12}$	-0.634	6.30e-05	-7.50e-04	-0.001	0.973	-0.068	-0.002	0.007	-0.017	2.02e-05
	(0.036)	(4.31e-05)	(1.23e-04)	(0.001)	(0.002)	(0.003)	(0.001)	(0.003)	(0.004)	(2.66e-06)
c_t	0.175	-0.002	0.004	-0.024	0.061	0.535	-0.033	-0.038	0.064	8.26e-05
	(0.325)	(6.03e-04)	(7.27e-04)	(0.010)	(0.018)	(0.024)	(0.009)	(0.021)	(0.029)	(1.69e-05)
$\Delta_c \ell_t$	10.137	5.29e-04	1.30e-05	0.011	0.023	-0.069	0.983	-0.163	0.660	3.05e-05
	(1.394)	(4.44e-04)	(2.60e-04)	(0.002)	(0.007)	(0.023)	(0.002)	(0.012)	(0.026)	(5.24e-06)
$\Delta_c y_t^{120,12}$	-0.324	-3.14e-05	-0.001	0.008	0.012	-0.058	-0.005	0.976	-0.054	2.23e-05
	(0.045)	(2.55e-04)	(2.17e-04)	(0.001)	(0.003)	(0.008)	(0.001)	(0.004)	(0.009)	(2.73e-06)
$\Delta_c \ell_t^{12,1}$	-0.004	-5.36e-04	-6.23e-05	-0.012	-0.025	0.069	0.019	0.188	0.437	-3.27e-05
	(0.004)	(4.48e-04)	(2.68e-04)	(0.002)	(0.008)	(0.023)	(0.002)	(0.012)	(0.026)	(5.47e-06)
e_t	$0.438^{\dagger}$	0.068	-0.315	-3.581	-0.095	0.308	-0.144	-1.757	1.271	0.996
	(0.411)	(0.028)	(0.038)	(0.453)	(0.936)	(0.811)	(0.207)	(0.683)	(0.631)	(3.77e-04)

Posterior mean and standard deviation for a two country model of the U.S. and Euro. We report unconditional means $\bar{\mu}_x^*, \bar{\mu}_x$, autocovariances matrices Φ_x^*, Φ_x, scale matrix Σ_x and the vector of loadings $\delta_{\widehat{c},x}$. The symbol † indicates that the unconditional means of the real exchange rates are not multiplied by 1200 (annualization and conversion to percent is not applicable).

Table 4: Posterior mean and stand. dev. of a two country model with the U.S.-Japan

x_t		Δs_t	$\Delta_c\pi_t$	ℓ_t	$y_t^{120,12}$	c_t	$\Delta_c\ell_t$	$\Delta_c y_t^{120,12}$	$\Delta_c\ell_t^{12,1}$	e_t
	$\bar{\mu}_x \times 1200$		Φ_x							
Δs_t	2.082	-0.011	-0.161	0.984	5.222	5.030	-1.396	-1.846	1.014	-0.014
	(0.374)	(0.050)	(0.187)	(0.885)	(2.950)	(3.368)	(1.691)	(3.222)	(2.968)	(0.010)
$\Delta_c\pi_t$	2.082	-0.002	0.162	-0.161	-0.131	0.973	0.111	-0.141	0.583	0.002
	(0.374)	(0.008)	(0.043)	(0.149)	(0.689)	(0.558)	(0.355)	(0.812)	(0.605)	(0.002)
ℓ_t	4.425	0.001	0.002	0.984	-0.040	-0.088	-0.059	-0.136	0.096	-2.66e-04
	(0.673)	(7.25e-04)	(0.005)	(0.013)	(0.048)	(0.053)	(0.026)	(0.054)	(0.046)	(1.48e-04)
$y_t^{120,12}$	1.650	-1.48e-05	-0.001	0.001	0.942	-0.021	-0.041	-0.021	-0.068	-1.53e-04
	(0.199)	(3.30e-04)	(0.002)	(0.006)	(0.027)	(0.028)	(0.014)	(0.032)	(0.025)	(7.45e-05)
c_t	0.652	1.76e-04	-0.010	0.057	-0.004	0.557	0.009	0.025	-0.008	9.10e-05
	(0.096)	(5.06e-04)	(0.003)	(0.010)	(0.038)	(0.041)	(0.021)	(0.045)	(0.036)	(1.11e-04)
$\Delta_c\ell_t$	2.547	0.002	-0.002	-0.047	-0.053	0.117	0.916	-0.283	0.807	-2.68e-04
	(0.401)	(8.74e-04)	(0.006)	(0.016)	(0.050)	(0.056)	(0.029)	(0.055)	(0.051)	(1.80e-04)
$\Delta_c y_t^{120,12}$	0.597	2.58e-04	-0.002	0.007	0.004	-0.017	-0.054	0.903	-0.103	-3.06e-04
	(0.235)	(3.56e-04)	(0.002)	(0.007)	(0.029)	(0.030)	(0.015)	(0.034)	(0.027)	(7.98e-05)
$\Delta_c\ell_t^{12,1}$	0.153	-8.42e-04	0.008	0.046	0.153	-0.090	0.003	0.078	0.258	9.11e-05
	(0.060)	(7.51e-04)	(0.005)	(0.014)	(0.045)	(0.050)	(0.026)	(0.050)	(0.046)	(1.56e-04)
e_t	$-5365.387^\dagger$	-0.009	-0.324	1.145	5.353	4.057	-1.508	-1.705	0.432	0.983
	(51.106)	(0.051)	(0.193)	(0.904)	(3.054)	(3.432)	(1.741)	(3.351)	(3.055)	(0.010)
x_t	$\delta_{\widehat{c},x}$		$\Sigma_x \times \sqrt{12} \times 100$							
Δs_t	0.002	10.966	0	0	0	0	0	0	0	
	(0.003)	(0.399)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	
$\Delta_c\pi_t$	0.007	-0.103	1.623	0	0	0	0	0	0	
	(0.004)	(0.084)	(0.059)	(—)	(—)	(—)	(—)	(—)	(—)	
ℓ_t	0.144	-0.009	0.008	0.153	0	0	0	0	0	
	(0.014)	(0.008)	(0.008)	(0.005)	(—)	(—)	(—)	(—)	(—)	
$y_t^{120,12}$	0.274	0.001	0.002	0.003	0.069	0	0	0	0	
	(0.048)	(0.003)	(0.003)	(0.003)	(0.002)	(—)	(—)	(—)	(—)	
c_t	-0.073	0.014	0.003	0.042	0.003	0.097	0	0	0	
	(0.080)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(—)	(—)	(—)	
$\Delta_c\ell_t$	-0.171	-0.009	0.009	0.066	0.002	0.016	0.170	0	0	
	(0.021)	(0.009)	(0.009)	(0.012)	(0.009)	(0.009)	(0.006)	(—)	(—)	
$\Delta_c y_t^{120,12}$	-0.459	0.003	-0.002	-0.002	0.057	0.005	-0.003	0.046	0	
	(0.063)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(—)	
$\Delta_c\ell_t^{12,1}$	0.478	-0.002	-0.003	-0.016	-0.016	-0.009	-0.139	-0.014	0.070	
	(0.167)	(0.008)	(0.008)	(0.011)	(0.008)	(0.008)	(0.006)	(0.004)	(0.003)	
e_t	-2.36e-04	11.070	-1.623	0	0	0	0	0	0	
	(9.18e-05)	(0.412)	(0.059)	(—)	(—)	(—)	(—)	(—)	(—)	
x_t	$\bar{\mu}_x^* \times 1200$		Φ_x^*							
Δs_t	7.678	0	0	0	0	0	1	0	0	0
	(1.630)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)	(—)
$\Delta_c\pi_t$	7.678	0.020	0.294	-2.182	-1.196	0.482	0.362	-1.191	-0.072	0.003
	(1.630)	(0.055)	(0.031)	(0.389)	(1.153)	(0.288)	(0.383)	(1.086)	(0.984)	(6.23e-04)
ℓ_t	25.052	1.98e-04	-0.002	0.999	0.044	-0.023	-0.006	-0.046	0.169	-7.41e-06
	(1.528)	(3.91e-04)	(9.14e-04)	(0.003)	(0.005)	(0.010)	(0.002)	(0.008)	(0.023)	(5.62e-06)
$y_t^{120,12}$	-0.644	1.10e-05	1.49e-04	1.02e-04	0.975	-0.069	-0.003	-0.002	-0.018	-3.30e-05
	(0.039)	(3.14e-05)	(8.94e-05)	(7.94e-04)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(3.60e-06)
c_t	0.939	-2.98e-04	-0.003	0.013	0.070	0.513	-0.018	-0.048	0.001	-1.96e-04
	(0.275)	(5.56e-04)	(6.95e-04)	(0.006)	(0.015)	(0.019)	(0.007)	(0.020)	(0.032)	(2.36e-05)
$\Delta_c\ell_t$	8.280	0.002	6.74e-05	0.002	-0.059	-0.004	0.973	-0.176	0.900	-4.50e-05
	(1.629)	(4.88e-04)	(3.89e-04)	(0.002)	(0.010)	(0.017)	(0.003)	(0.010)	(0.030)	(9.99e-06)
$\Delta_c y_t^{120,12}$	-0.281	1.36e-04	5.12e-04	0.005	0.002	-0.080	-0.011	0.971	-0.046	-4.57e-05
	(0.055)	(2.04e-04)	(3.06e-04)	(0.001)	(0.004)	(0.006)	(0.002)	(0.005)	(0.013)	(5.03e-06)
$\Delta_c\ell_t^{12,1}$	-5.77e-04	-0.002	-4.38e-05	-0.002	0.065	0.001	0.029	0.195	0.195	4.78e-05
	(0.005)	(4.92e-04)	(3.95e-04)	(0.002)	(0.011)	(0.017)	(0.003)	(0.010)	(0.031)	(1.04e-05)
e_t	$-5891.507^\dagger$	-0.020	-0.294	2.182	1.196	-0.482	0.638	1.191	0.072	0.997
	(494.508)	(0.055)	(0.031)	(0.389)	(1.153)	(0.288)	(0.383)	(1.086)	(0.984)	(6.23e-04)

Posterior mean and standard deviation for a two country model of the U.S. and Japan. We report unconditional means $\bar{\mu}_x^*, \bar{\mu}_x$, autocovariances matrices Φ_x^*, Φ_x, scale matrix Σ_x and the vector of loadings $\delta_{\widehat{c},x}$. The symbol † indicates that the unconditional means of the real exchange rates are not multiplied by 1200 (annualization and conversion to percent is not applicable).

Table 5: Pricing errors across countries

	U.K.	CAN	EURO	JPN
$\ell_t^1 - \ell_t^3 + \widehat{\ell}^3$	0.045	0.049	0.048	0.058
$\ell_t^1 - \ell_t^6 + \widehat{\ell}^6$	0.035	0.035	0.038	0.040
y_t^3	0.061	0.049	0.052	0.045
y_t^{12}	0.113	0.109	0.097	0.098
y_t^{24}	0.150	0.146	0.132	0.139
y_t^{36}	0.160	0.157	0.142	0.151
y_t^{48}	0.158	0.155	0.140	0.149
y_t^{60}	0.150	0.148	0.132	0.141
y_t^{84}	0.131	0.129	0.115	0.120
$\widehat{y}_t^3$	0.109	-	-	-
$\widehat{y}_t^{24}$	0.155	0.154	0.146	0.130
$\widehat{y}_t^{36}$	0.165	0.155	0.152	0.142
$\widehat{y}_t^{48}$	0.163	0.150	0.150	0.151
$\widehat{y}_t^{60}$	0.155	0.145	0.145	0.154
$\widehat{y}_t^{84}$	0.138	0.138	0.134	0.151
$\widehat{y}_t^{120}$	0.118	0.128	0.122	0.125

Posterior mean estimates of the pricing errors in annualized percentage points, $100\sqrt{\mathrm{diag}\left(\Sigma_y \Sigma_y'\right) \times 12}$, *for the U.K., Canada, Euro, and Japan. These are reported for yields of different maturity that enter* $Y_t^{(2)}$.

Table 6: Estimates of the speed of mean reversion coefficients

	U.K. α_f	U.K. α_f^*	CAN α_f	CAN α_f^*	EURO α_f	EURO α_f^*	JPN α_f	JPN α_f^*
Δs_t	-0.189	0	-0.094	0	-0.196	0	-0.082	0
	(0.053)	(—)	(0.050)	(—)	(0.056)	(—)	(0.054)	(—)
$\Delta_c \pi_t$	0.175	0.356	0.097	0.245	0.115	0.140	0.082	0.122
	(0.066)	(0.054)	(0.056)	(0.031)	(0.066)	(0.015)	(0.069)	(0.024)
ℓ_t	0.006	0.014	0.004	0.002	0.003	-0.002	-0.018	-5.13e-04
	(0.011)	(0.003)	(0.009)	(0.001)	(0.011)	(6.53e-04)	(0.010)	(3.89e-04)
$y_t^{120,12}$	-0.039	-0.010	-0.005	-9.46e-04	-0.073	0.003	-0.030	-0.006
	(0.014)	(0.001)	(0.012)	(4.52e-04)	(0.014)	(4.45e-04)	(0.015)	(7.05e-04)
c_t	0.098	0.021	0.035	0.027	0.099	0.025	0.032	-0.070
	(0.040)	(0.021)	(0.035)	(0.008)	(0.041)	(0.005)	(0.039)	(0.008)
$\Delta_c \ell_t$	0.064	0.007	-0.011	-0.002	-0.008	0.002	-0.026	-0.004
	(0.018)	(0.002)	(0.018)	(0.001)	(0.018)	(3.91e-04)	(0.017)	(9.68e-04)
$\Delta_c y_t^{120,12}$	-0.030	0.002	0.024	-9.42e-04	-0.039	0.003	-0.050	-0.007
	(0.017)	(0.002)	(0.018)	(0.001)	(0.015)	(3.84e-04)	(0.013)	(8.19e-04)
$\Delta_c \ell_t^{12,1}$	-0.120	-0.023	0.009	0.007	-0.003	-0.009	0.030	0.016
	(0.043)	(0.006)	(0.044)	(0.003)	(0.052)	(0.001)	(0.052)	(0.003)

Posterior mean and standard deviation estimates of α and α^* for the U.K., Canada, Euro, and Japan. The factors have been re-scaled to have unit variance.