Currency Risk in the Long Run*

Pasquale Della Corte
Imperial College London
& CEPR

Can Gao
SAFE

Daniel P.A. Preve
SMU

Giorgio Valente
HKMA

January 2022

*The authors would like to thank Juan Antolin Diaz, Gino Cenedese, Max Croce, Bernard Dumas, Ľuboš Pástor, Fabio Trojani, conference participants at the 2018 Workshop on Bayesian Methods in Finance in Singapore, 2018 Belgrade Symposium in Economics and Finance, 2021 Asset Pricing Conference by Long-Term Investors@UniTo as well as seminar participants at the Bank of England and Fulcrum Asset Management for their helpful and constructive comments. The views expressed in this paper are those of the authors and are not necessarily those of the Bank of England or the Hong Kong Monetary Authority. All remaining errors are our own. Pasquale Della Corte is with Imperial College Business School, Imperial College London and the Centre for Economic Policy Research (CEPR), London. Email: p.dellacorte@imperial.ac.uk. Can Gao is with Leibniz Institute for Financial Research, SAFE, Frankfurt. Email: gao@safe-frankfurt.de. Daniel P.A. Preve is with the School of Economics, Singapore Management University. Email: dpreve@smu.edu.sg. Giorgio Valente is with Hong Kong Institute for Monetary Research, Hong Kong Monetary Authority, Hong Kong. Email: valentegiorgio@icloud.com.

Currency Risk in the Long Run

Abstract

We study the long-horizon risk profile of a currency strategy, whereby a US investor earns excess returns by entering in an unhedged long position in a foreign long-term bond funded at the domestic risk-free rate. After showing the drivers of the strategy returns, we derive and estimate their long-horizon predictive variance using data on long-term bonds denominated in major currencies over the past two centuries. We find that the long-horizon risk of such strategies increases with the investment horizon and that it is mainly driven by the uncertainty associated with the predictions of future returns originating from interest rate differentials and exchange rate returns.

Keywords: Currency risk, Long-term bonds, Predictability, Long-term investments.

JEL Classification: F31, G12, G15.

"Demand for longer-dated higher-yielding cash flows is very, very present."

Scott Thiel, Head of Global Bonds at BlackRock, 2017.

"Insatiable demand for long bonds isn't short term."

The Wall Street Journal, 2017.

1 Introduction

Recent financial turmoils, like the ones associated with the collapse of Lehman Brothers and the coronavirus outbreak, have been followed by unconventional monetary policy interventions aiming at offsetting the negative impact of economic recessions on employment and growth. These monetary stimuli have produced negative short-term interest rates in most developed countries and pushed investors in search for higher yields to consider long-term bonds around the world as an alternative asset class (e.g., Allen, 2017). Regulatory changes have also contributed to increase market participants' appetite for long-dated debt instruments. For example, international banks have increased their demand for sovereign bonds to meet the liquidity requirements arising from the post-crisis financial regulation, as they carry little regulatory risk (e.g., Acharya and Steffen, 2015). Pension funds and other institutional investors, moreover, have also increased their exposure to long-duration assets to better match their liabilities and hedge interest rate risk (e.g., IMF, 2019).

Furthermore, divergent monetary policy stances have also created an incentive for international investors to increase their reliance on short-term dollar funding, thus increasing their exposure to both interest rate and foreign exchange rate (e.g., Maggiori, Neiman and Schreger, 2020; Cenedese, Della Corte and Wang, 2021). Global bond and foreign exchange markets have thus become increasingly more integrated and unexpected movements in short-term US interest rates can affect those investors that have exposure to both foreign exchange rate and foreign long-term bonds (e.g., Maggiori, 2017; Greenwood, Hanson, Stein and Sunderam, 2019).[1] A related literature studies the behavior of a carry trade strategy implemented with long-term foreign bonds and finds a near zero average excess return, which conceals a large negative profitability before the crisis followed by a sharp positive performance after the global financial crisis (Lustig, Stathopoulos and Verdelhan, 2019; Andrews, Colacito, Croce and Gavazzoni, 2021).

In this paper, we explore the long-horizon variance risk faced by a US investor holding an unhedged position in a long-term foreign bond funded at the domestic risk-free rate. While early academic studies suggest that investors would naturally hold short-term bonds and demand a premium to hold long-term bonds (e.g., Keynes, 1930; Hicks, 1946), a more recent literature argues that long-term bonds are more appropriate for investors that value stability of income (Modigliani and Sutch, 1966) and lower risk (Campbell and Viceira, 2002; Viceira and Wang, 2018). Although some studies have attempted to provide empirical evidence to these issues, to date little is known about the risk profile of a strategy that relies on short-term dollar funding and buys long-term foreign currency-denominated bonds. We first show that the excess returns from this strategy are due to three main components: The bond excess returns in local currency, the short-term real interest differential between the foreign and domestic currency and the bilateral real exchange rate return.

We then adopt the predictive variance of the strategy returns as a notion of risk in our analysis. In fact, the predictive variance of returns is what really matters to investors, especially for long-horizon portfolio decisions, as it requires the identification and estimation of suitable predictive models for asset returns (Pástor and Stambaugh, 2012; Avramov, Cederburg and Lučivjanská, 2018). Investors construct their expectations of future returns using only information available at the time the forecast is made. However, they do not know the true

[1] During the financial crisis in 2008, many developing economies received large waves of capital inflows, which sharply reversed following the decision of the US Federal Reserve to start withdrawing monetary stimulus. This lead Morgan Stanley to coin the term "fragile five" to refer a group of developing economies that were suffering sharp outflows of foreign capital, exchange rate depreciation, and difficulty to roll-over their debts.

data generating process of future returns and, by relying on potentially misspecified empirical models, observable predictors may deliver imperfect forecasts. Hence, the investors' predictive variance of returns differ substantially from the true variance in that the former encompasses a range of uncertainties which are absent in the latter. These uncertainties are important for long-term investors as they are likely to offset the effects of mean reversion in returns for longer investment horizons, even in the presence of return predictability.[2]

We carry out an empirical investigation exploring the long-term predictive variances of returns from investments in long-term bonds denominated in major currencies over the past two centuries. In the spirit of Pástor and Stambaugh (2009), we estimate these long-term predictive variances in an environment with imperfect predictability, by allowing unobserved predictors to join a set of observable predictors to help forecast the strategy returns at different horizons. As the variables to be forecast in our setting relate to interest rates and exchange rates, we conjecture that the unobserved predictors in our framework can be potentially associated with changes in monetary and exchange rate regimes that have occurred over the past two centuries and are not already captured by the set of observable predictors. Furthermore, we derive in closed form a range of uncertainties that affect the predictive variance for long horizons, in addition to the mean reversion component due to predictability of returns, building upon and extending the theoretical frameworks proposed by Pástor and Stambaugh (2012) and Avramov et al. (2018).

The estimations lead to a host of interesting results: First, over the full sample period and across all countries, the predictive variance of the bond investment strategy is found to be increasing with the investment horizon and this is mainly due to a growing predictive variance for both short-term interest rate differentials and real exchange rate returns. Overall, the predictive variance of real exchange rate returns exhibits the largest long-horizon value, followed by that of interest rate differentials, while the predictive variance of bond excess returns in foreign currency does not vary much across investment horizons. The predictive

[2]The effect of these uncertainties on the predictive variance may be even stronger if returns are only partially predictable.

covariances between the three components of the strategy returns are smaller in size and negligible, with the exception of the covariance between interest rate differentials and real exchange rate returns which is consistently negative across currency pairs and decreasing over the investment horizon. This finding suggests that predictive co-movement between bond excess returns in foreign currency and interest rates and exchange rate returns are less important in determining the long-term risk profile of the strategy. The results also suggest that the predictive co-movements between interest rate differentials and real exchange rate returns are important in the long-run as they tend to reduce the overall expected risk of the strategy, especially at longer horizons. Second, after decomposing the predictive variance of the strategy returns into its key constituents, we observe that in all cases the uncertainty about future returns unambiguously plays the leading role. All other components, especially the one associated with mean reversion due to return predictability, are negligible in size and do not provide any improvement in the risk profile of the strategy at longer horizons. Put differently, when bond return, interest rate and foreign exchange rate predictability is taken seriously into account, the notion that bond returns are less volatile in the long-run does not apply. The range of uncertainties that affect the predictive variance more than offset any potential benefit originating from mean reversion in returns.

Upon further exploration, across all of the currency pairs investigated, the uncertainty about future returns is mainly due to the component of the predictive variance pertaining to the unobserved predictors. The uncertainty associated with the expected future values of the observable predictors is non-negligible but substantially smaller than the one documented for unobserved predictors. Furthermore, as the uncertainty about future returns originating from interest rate differentials and exchange rate returns is fairly similar, the shape of the predictive variance over longer horizons can be interpreted as spurring from changes in monetary and exchange rate regimes that are not captured by the set of observable predictors.

We do not find any tangible effects of mean reversion in the long-horizon predictive variance as, for most of the currencies considered, the expected negative impact of the mean reversion component on the predictive variance (only found for unobserved predictors) is completely offset by a similar but positive impact of the mean reversion component originating from observable predictors. As the overall effect cancels out, the shape of the strategy's predictive variance is essentially due to the interplay between the various sources of uncertainties and therefore increases with the investment horizon.

Our study builds on the findings of various strands of literature: First and foremost, a vast body of research that has extensively explored the issue of whether bond returns and yields are predictable over various investment horizons. In this body, several studies have found evidence of predictability for US bond yields and returns provided by, but not limited to, interest forward rates, macroeconomic fundamentals and principal components of bond yields (Fama and Bliss, 1987; Cochrane and Piazzesi, 2005; Ludvigson and Ng, 2009; Joslin, Priebsch and Singleton, 2014). Evidence of predictability for non-US bond returns is reported in some studies, but it is overall less pervasive than that found for US bond returns (Ilmanen, 1995). As we are concerned with the risk of investing in foreign long-term bonds, the predictability of FX returns is also important. This literature, however, has not reached a consensus as to whether, and to which extent, FX returns are at all predictable (Meese and Rogoff, 1983; Engel and West, 2005). In our study, we allow for predictability in all components of the strategy returns but we also take into account the possibility that such predictability is less than perfect, à la Pástor and Stambaugh (2009), in order to encompass the different levels of predictability that have been documented in the existing literature. The second body of research we build on relates to the measurement of risk associated with portfolio investments over long horizons, and the implications for asset allocation. In addition to the pioneering works by Samuelson (1969) and Merton (1969), who show that investors should choose the same asset allocation regardless of investment horizon whenever asset returns are unpredictable, our empirical analysis builds upon the results of Siegel (1992, 2008), Barberis (2000) and Campbell and Viceira (2002, 2005). These studies show that, in the presence of return predictability, the perceived variability of asset returns is lower for longer horizons because of the effect mean reversion of expected returns has on the long-horizon variance. The two studies that are closest to our empirical investigation are the ones by Pástor and

Stambaugh (2012) and Avramov et al. (2018), who show that asset returns are more volatile over longer horizons if the predictive variance of returns is used as the main notion of long-horizon risk. This novel result is due to the presence of an assortment of uncertainties that are explicitly included in the framework for the predictive variance, but not for the true variance. We improve upon these works in several important ways: We first show how the long-horizon predictive covariances associated with a predictive system that includes both observable and unobserved predictors can be decomposed into five main components, with accompanying closed-form expressions. Thus, in contrast to Pástor and Stambaugh (2012) and Avramov et al. (2018), who focus on the long-horizon predictive variance of a single asset, our focus is on obtaining an informative decomposition for the long-horizon predictive covariance of multiple assets. We then apply this framework to a multiple asset case, allowing for imperfect predictability. This permits us to gain further insight into the long-horizon predictive variance of the strategy returns and to directly link the main sources of uncertainty, including mean reversion in expected returns, to the estimated parameters of the predictive system.

Our study is structured as follows: Section 2 presents our framework for deriving and computing long-horizon predictive covariances in the presence of imperfect predictability, and discusses the theoretical findings. Section 3 shows the components of the strategy returns and introduces their long-horizon predictive variance. It also describes the long-span data used in the empirical investigation and reports some preliminary statistics. Section 4 reports the main results and discuss a number of robustness checks, respectively, whereas Section 5 concludes. A separate Internet Appendix describes the Bayesian estimation and presents the derivation of the decomposition for the long-horizon predictive covariance.

2 Framework

This section presents our framework for long-horizon predictive variances and covariances. It builds upon and extends the work of Pástor and Stambaugh (2012) and Avramov et al. (2018). We start with a parsimonious predictive system that includes both observable and unobserved predictors. We then show how the long-horizon predictive covariances associated with this system can be decomposed into five main components, with accompanying closed-form expressions. Section 3 then uses this framework to empirically examine the predictive return variance over long horizons of a strategy that buys a long-term bond in foreign currency while borrowing at the short-term domestic interest rate. The excess return on this strategy can be seen as the sum of a bond excess return in local currency, a real interest rate differential, and real exchange rate return. Hence its predictive return variance, by construction, comprises both variance and covariance terms. Since variance is a special case of covariance, we only present the framework for covariances here.

2.1 A Simple Currency Strategy

Similar to Andrews et al. (2021), we consider a simple strategy where a US investor buys a long-term bond in foreign currency and sells a short-term bond in local currency. The one-month excess return on this strategy, rx_{t+1}, can be described as

$$rx_{t+1} = y^{\star}_{t+1} + \Delta s_{t+1} - i_{t+1}, \tag{1}$$

where $y^{\star}_{t+1}$ is the one-month return on a constant maturity long-term bond denominated in foreign currency, Δs_{t+1} is the one-month nominal exchange rate return, and i_{t+1} is the one-month return on a short-term bond denominated in domestic currency.

We can equivalently rewrite the excess return using real quantities, i.e., we add and subtract the one-month return on a short-term bond denominated in foreign currency $i^{\star}_{t+1}$, the one-month domestic inflation rate ρ_{t+1}, and the one-month foreign inflation rate $\rho^{\star}_{t+1}$ as

$$rx_{t+1} = \underbrace{y^{\star}_{t+1} - i^{\star}_{t+1}}_{\substack{\textit{foreign bond} \\ \textit{excess return}}} + \underbrace{(i^{\star}_{t+1} - \rho^{\star}_{t+1}) - (i_{t+1} - \rho_{t+1})}_{\substack{\textit{real interest} \\ \textit{rate differential}}} + \underbrace{\Delta s_{t+1} + \rho^{\star}_{t+1} - \rho_{t+1}}_{\substack{\textit{real exchange} \\ \textit{rate retun}}} \tag{2}$$

or simply as the sum of three return components

$$rx_{t+1} = r_{1,t+1} + r_{2,t+1} + r_{3,t+1}, \tag{3}$$

where $r_{1,t+1} = y^{\star}_{t+1} - i^{\star}_{t+1}$ is the foreign bond excess return in local currency, $r_{2,t+1} = (i^{\star}_{t+1} - \rho^{\star}_{t+1}) - (i_{t+1} - \rho_{t+1})$ is the real interest rate differential between the foreign and the domestic country, and $r_{3,t+1} = \Delta s_{t+1} + \rho^{\star}_{t+1} - \rho_{t+1}$ is the real exchange rate return. All variables are defined between times t and $t+1$ and the asterisk refers to foreign quantities.

2.2 Imperfect Predictability

Each return component in Equation (3) can be written as

$$r_{s,t+1} = \mu_{s,t} + u_{s,t+1} \tag{4}$$

where $\mu_{s,t}$ denotes the expected return conditional on *all* information available at time t, $u_{s,t+1}$ is the unexpected return with zero mean and constant variance, and $s \in \{1, 2, 3\}$. The expected return $\mu_{s,t}$ is often defined as a linear combination of an observable predictor $x_{s,t}$ (or a set of observable predictors) such that $\mu_{s,t} = a_s + b_s x_{s,t}$. As noted by Pástor and Stambaugh (2009), this approach is likely to understate the uncertainty faced by an investor assessing the variance of future returns. This happens as the true expected return $\mu_{s,t}$ reflects more information than what we assume an investor can observe, i.e., the history of $r_{s,t}$ and $x_{s,t}$. Put differently, any observable predictor $x_{s,t}$ is likely to be *imperfect* and unable to deliver the true expected return, i.e., $\mu_{s,t} \neq a_s + b_s x_{s,t}$. On the basis of this arguments, we take the presence of predictor imperfection into account by first defining the expected return as

$$\mu_{s,t} = a_s + b_s x_{s,t} + \pi_{s,t}, \tag{5}$$

where $\pi_{s,t}$ denotes the unobserved predictor, and then considering a state-space model in which $r_{s,t}$, $x_{s,t}$, and $\pi_{s,t}$ follow a first-order vector autoregression with coefficients restricted such that $\mu_{s,t}$ is the mean of $r_{s,t}$.

2.3 Long-Horizon Variance and Parameter Uncertainty

We study the predictive variance over long horizons of rx_{t+1} and how the shape of the variance curve is affected by its underlying components. Unlike the (ex post) realized variance, which implicitly assumes full knowledge of the data generating process, the (ex ante) predictive variance only conditions on information available to an investor and incorporates parameter uncertainty to make forward-looking predictions.

Define the k-period return from period $T+1$ through period $T+k$ as

$$r_{s,T}^{k} = \sum_{\ell=1}^{k} r_{s,T+\ell}$$

and assess the predictive variance of our excess return as

$$Var\left(rx_T^k \mid D_T\right) = \sum_{i=1}^{3}\sum_{j=1}^{3} Cov\left(r_{i,T}^k, r_{j,T}^k \mid D_T\right), \tag{6}$$

where $rx_T^k = r_{1,T}^k + r_{2,T}^k + r_{3,T}^k$ is the k-period excess return and D_T denotes a subset of all information available at time T. In our empirical analysis, D_T comprises the full history of returns $r_{s,t}$ and observable predictors $x_{s,t}$ that an investor employs to forecast returns. It does not contain, however, any information on the unobservable predictor $\pi_{s,t}$ and the vector of parameters ϕ governing the joint dynamics of returns and predictors. The elements of ϕ are considered as random, given that they are unknown to an investor.

We focus on $Var\left(rx_T^k \mid D_T\right)$, i.e., the predictive variance of rx_T^k given the information set D_T available to an investor at time T, which comprises both variance and covariance terms. As the variance is a special case of the covariance, we present our framework for $Cov\left(r_{i,T}^k, r_{j,T}^k \mid D_T\right)$, i.e., the predictive covariance between $r_{i,T}^k$ and $r_{j,T}^k$ given the information set D_T available to an investor at time T. Given this framework, we can first calculate all predictive variances and covariances implied by Equation (6), and then the predictive variance of rx_T^k.

In our setting, an investor is uncertain about π_T and ϕ and it is helpful to decompose the predictive covariance between $r_{i,T}^k$ and $r_{j,T}^k$ as

$$\begin{aligned} Cov\left(r_{i,T}^k, r_{j,T}^k \mid D_T\right) = & \; E\left[Cov\left(r_{i,T}^k, r_{j,T}^k \mid \pi_T, \phi, D_T\right) \mid D_T\right] \\ & + Cov\left[E\left(r_{i,T}^k \mid \pi_T, \phi, D_T\right), E\left(r_{j,T}^k \mid \pi_T, \phi, D_T\right) \mid D_T\right]. \end{aligned} \tag{7}$$

The first term of this decomposition is the expectation of the conditional covariance of k-period returns. While investors with a knowledge of the true values of π_T and ϕ only care about the conditional covariance, investors who are uncertain about π_T and ϕ care about its expectation and also also account for the covariance of the conditional expected k-period returns, the second term in Equation (7). As a result, the perceived covariance can be substantially higher, in absolute terms, at long horizons. To assess these effects, we first need to define a predictive system that describes the evolution of our return components and then dissect each of these components using closed-form solutions.

2.4 Predictive System

We model the return components underlying the excess return rx_t using a parsimonious predictive system in vector form defined as

$$r_{t+1} = a + bx_t + \pi_t + u_{t+1} \tag{8}$$

$$x_{t+1} = \theta + \gamma x_t + v_{t+1} \tag{9}$$

$$\pi_{t+1} = \delta \pi_t + \eta_{t+1}, \tag{10}$$

where $r_t = [r_{1,t}, \ldots, r_{3,t}]'$ is a vector of returns, $x_t = [x_{1,t}, \ldots, x_{3,t}]'$ is a vector of observable predictors, and $\pi_t = [\pi_{1,t}, \ldots, \pi_{3,t}]'$ is a vector of unobserved predictors. All vectors r_t, x_t and π_t have a size 3×1. Also, a and θ denote 3×1 vectors of intercepts whereas b, γ and δ

represent 3×3 diagonal matrices of slope coefficients. For the main diagonal elements of γ and δ, we assume that $-1 < \gamma_s < 1$ and $0 < \delta_s < 1$ for $s \in \{1, 2, 3\}$.

In our empirical investigation, the vectors containing the shocks of the system are assumed to be independent and identically normally distributed over time

$$\begin{bmatrix} u_t \\ v_t \\ \eta_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma'_{vu} & \Sigma'_{\eta u} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma'_{\eta v} \\ \Sigma_{\eta u} & \Sigma_{\eta v} & \Sigma_{\eta\eta} \end{bmatrix} \right), \tag{11}$$

where Σ_{uu} is the covariance matrix of the unexpected returns u_t, Σ_{vv} is the covariance matrix of the observable predictors' shocks v_t, $\Sigma_{\eta\eta}$ is the covariance matrix of the unobserved predictors' shocks η_t, and the off-diagonal entries denote the cross-equations covariance matrices.[3] All submatrices in (11) are of size 3×3. We refer to each element of, say, $\Sigma_{\eta u}$ as $\sigma_{\eta_i u_j}$ for $i, j \in \{1, 2, 3\}$.

The predictive system presented above can be seen as a reduced-form model that is consistent with a broad range of economic models, rational or behavioral, in which the expected return changes over time in a persistent fashion. In particular, each return $r_{s,t+1}$ is a linear function of a lagged observable predictor $x_{s,t}$ and a lagged unobserved predictor $\pi_{s,t}$. Each observable predictor $x_{s,t}$, in turn, follows a first-order autoregressive process, an assumption routinely used in the predictability literature (e.g., Stambaugh, 1999). A special case arises when the coefficients b_s (the main diagonal elements of b) are zero, observable predictors are then absent and returns are only driven by unobserved predictors. Finally, in Equation (10), we postulate a persistent driftless process for each unobserved predictor $\pi_{s,t}$. This equation is particularly useful to study the role of predictor imperfection. To see this, notice that unobserved predictors are absent in the limit as the main diagonal elements of $\Sigma_{\eta\eta}$ (the $\sigma^2_{\eta_s}$) tend to zero. Equation (8) then reduces to a standard predictive regression used in a wide range of empirical applications. We will conduct our empirical analysis using monthly

[3]Our theoretical results presented in sections 2.5-2.6, however, hold under the considerably less restrictive assumption of a zero-mean vector white noise process for the shocks of the predictive system.

observations between January 1800 and June 2017 for 10 major economies, and compute posterior distributions for the parameters of our predictive system using the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B.

2.5 Conditional Covariance

The conditional covariance $Cov\left(r_{i,T}^{k}, r_{j,T}^{k} \mid \pi_T, \phi, D_T\right)$ on the right-hand side of Equation (7) implies that the investor knows both π_T and ϕ and, hence, takes neither uncertainty about the current expected return nor parameter uncertainty into account. Assuming that equations (8)–(11) hold, this important conditional covariance comprises three main sources of uncertainty, denoted by S_1 through S_3, as

$$Cov\left(r_{i,T}^{k}, r_{j,T}^{k} \mid \pi_T, \phi, D_T\right) = S_1 + S_2 + S_3, \tag{12}$$

where

$$S_1 = k\sigma_{u_i}\sigma_{u_j}\rho_{u_i u_j} \tag{13}$$

$$\begin{aligned} S_2 &= k\sigma_{u_i}\sigma_{u_j}\left[b_i\bar{e}_i\rho_{v_i u_j}A_{\gamma_i}(k) + b_j\bar{e}_j\rho_{u_i v_j}A_{\gamma_j}(k)\right] \\ &\quad + k\sigma_{u_i}\sigma_{u_j}\left[\bar{d}_i\rho_{\eta_i u_j}A_{\delta_i}(k) + \bar{d}_j\rho_{u_i\eta_j}A_{\delta_j}(k)\right] \end{aligned} \tag{14}$$

$$\begin{aligned} S_3 &= k\sigma_{u_i}\sigma_{u_j}b_ib_j\bar{e}_i\bar{e}_j\rho_{v_iv_j}B_{\gamma_i,\gamma_j}(k) + k\sigma_{u_i}\sigma_{u_j}\bar{d}_i\bar{d}_j\rho_{\eta_i\eta_j}B_{\delta_i,\delta_j}(k) \\ &\quad + k\sigma_{u_i}\sigma_{u_j}\left[b_i\bar{e}_i\bar{d}_j\rho_{v_i\eta_j}B_{\gamma_i,\delta_j}(k) + b_j\bar{d}_i\bar{e}_j\rho_{\eta_i v_j}B_{\delta_i,\gamma_j}(k)\right], \end{aligned} \tag{15}$$

and

$$\begin{aligned} A_{\chi_s}(k) &= 1 + \frac{1}{k}\left(-1 - \chi_s\frac{1-\chi_s^{k-1}}{1-\chi_s}\right) \\ B_{\chi_i,\psi_j}(k) &= 1 + \frac{1}{k}\left(-1 - \chi_i\frac{1-\chi_i^{k-1}}{1-\chi_i} - \psi_j\frac{1-\psi_j^{k-1}}{1-\psi_j} + \chi_i\psi_j\frac{1-\chi_i^{k-1}\psi_j^{k-1}}{1-\chi_i\psi_j}\right) \end{aligned}$$

$$\bar{d}_s = \left(\frac{1+\delta_s}{1-\delta_s} \frac{R_s^2}{1-R_s^2} \frac{\sigma_{\pi_s}^2}{\sigma_{r_s}^2 - \sigma_{u_s}^2} \right)^{1/2}$$

$$\bar{e}_s = \left(\frac{1+\gamma_s}{1-\gamma_s} \frac{R_s^2}{1-R_s^2} \frac{\sigma_{x_s}^2}{\sigma_{r_s}^2 - \sigma_{u_s}^2} \right)^{1/2},$$

for $\chi, \psi = \gamma, \delta$ and $s = i, j$ and $R_s^2 = (b_s^2 \sigma_{x_s}^2 + \sigma_{\pi_s}^2 + 2b_s \sigma_{x_s \pi_s})/\sigma_{r_s}^2$ is the variance of $\mu_{s,t}$ to the variance of $r_{s,t+1}$ based on Equation (8). We report the corresponding derivation in the Internet Appendix A.

The first source of uncertainty, S_1 in Equation (13), can be interpreted as the uncertainty arising from $i.i.d.$ shocks and contributes the same per-period covariance at all horizons. The two terms of the second source of uncertainty, S_2 in Equation (14), reflect the correlation between unexpected returns and shocks to expected future returns. The first term captures the correlation between unexpected returns and shocks to observable predictors through $\rho_{v_i u_j}$ and $\rho_{u_i v_j}$, the second term represents the correlation between unexpected returns and shocks to unobserved predictors by means of $\rho_{\eta_i u_j}$ and $\rho_{u_i \eta_j}$. When these correlations are negative, the terms on the right-hand side of Equation (14) will contribute negatively to long-horizon covariance given typical parameter estimates, thus reflecting the mean-reverting dynamics of expected returns. The literature on stock returns (e.g., Campbell, 1991; Campbell, Chan and Viceira, 2003), for example, finds a negative correlation between unexpected returns and shocks to expected returns and concludes that stocks have lower per-period variance and are less risky for long-horizon investors.

The third source of uncertainty, S_3 in Equation (15), reflects uncertainty about future expected returns and it contributes positively to long-horizon covariance given typical parameter estimates. Even with perfect information on the parameters of the predictive system, and on the current values of its predictors, an investor is still uncertain about future expected returns (or, equivalently, about future values of the observable and unobserved predictors) in each period. This uncertainty produces additional predictive covariance that is often ignored in the literature. Specifically, the first term of S_3 captures uncertainty about future values of the observable predictors while the second term reflects uncertainty about future values of the unobserved predictors. The last term, instead, captures joint uncertainty about future values of the observable and unobserved predictors.

When returns are unpredictable, S_1 is the only non-zero source of uncertainty in Equation (12). When returns are predictable (either via observable predictors, unobserved predictors, or both), S_1 is still the only non-zero source when $k = 1$, since both $A_{\chi_s}(1)$ and $B_{\chi_i,\psi_j}(1)$ are zero. As k increases, however, the terms of S_2 and S_3 involving $A_{\chi_s}(k)$ and $B_{\chi_i,\psi_j}(k)$ become increasingly important, since both functions are strictly increasing from zero to one as k tends from one to infinity, and all three sources in Equation (12) play a role. When asset s has no observable predictor ($b_s = 0$), $R_s^2 = \sigma_{\pi_s}^2/\sigma_{r_s}^2$, and the first term in Equation (14) as well as the first and third terms in Equation (15) vanish.

Finally, when $i = j$ and there are no observable predictors the conditional covariance in Equation (12) becomes the conditional variance

$$Var\left(r_{i,T}^k \mid \pi_T, \phi, D_T\right) = k\sigma_{u_i}^2\left[1 + 2\bar{d}_i\rho_{u_i\eta_i}A_{\delta_i}(k) + \bar{d}_i^2 B_{\delta_i,\delta_i}(k)\right],$$

which coincides with the expression for the conditional variance presented in Pástor and Stambaugh (2012, p. 438).

2.6 Components of Predictive Covariance

The long-horizon predictive covariance $Cov\left(r_{i,T}^k, r_{j,T}^k \mid D_T\right)$ in Equation (7) comprises two terms. The first term $E\left[Cov\left(r_{i,T}^k, r_{j,T}^k \mid \pi_T, \phi, D_T\right) \mid D_T\right]$ is the expectation of a conditional covariance of k-period returns, which corresponds to the sum of the on D_T conditional expectations of S_1–S_3, the three sources of uncertainty in Equation (12). The second term $Cov\left[E\left(r_{i,T}^k \mid \pi_T, \phi, D_T\right), E\left(r_{j,T}^k \mid \pi_T, \phi, D_T\right) \mid D_T\right]$ on the right-hand side of Equation (7) is the covariance of the true conditional expected returns given the investor's information set D_T. In Internet Appendix A, we decompose this covariance into two components. The first component reflects uncertainty about the current π_T (or predictor imperfection), and the second reflects uncertainty about ϕ (or estimation risk).

From the perspective of an investor, an important consideration is whether the predictive covariance of $r^k_{i,T}$ and $r^k_{j,T}$, the k-period returns of assets i and j starting at time T, increases or decreases with the investment horizon k and, generally, what the shape of the covariance curve is.

Analogous to Pástor and Stambaugh (2012), let $c_{s,T}$ denote the on ϕ and D_T conditional mean of the unobserved predictor $\pi_{s,T}$ and let $q_{ij,T}$ denote the conditional covariance between the unobserved predictors $\pi_{i,T}$ and $\pi_{j,T}$. That is,

$$c_{s,T} = E(\pi_{s,T} \mid \phi, D_T)$$
$$q_{ij,T} = Cov(\pi_{i,T},\, \pi_{j,T} \mid \phi, D_T).$$

To gain further insight into the long-horizon predictive covariance, we can then express the right-hand side of Equation (7) as the sum of five components. These components, denoted by C_1 through C_5 in Equation (16) below, are discussed separately in sections 2.6.1–2.6.4.

$$\begin{aligned} &Cov\left(r^k_{i,T}, r^k_{j,T} \mid D_T\right) \\ &= \underbrace{C_1}_{i.i.d.\ uncertainty} + \underbrace{C_2}_{mean\ reversion} \\ &+ \underbrace{C_3}_{future\ \mu_s\ uncertainty} + \underbrace{C_4}_{current\ \pi_s\ uncertainty} + \underbrace{C_5}_{estimation\ risk}. \end{aligned} \tag{16}$$

The derivation of Equation (16) is presented in Internet Appendix A.

2.6.1 i.i.d. uncertainty

The first component of the long-horizon predictive covariance in Equation (16), labeled *i.i.d. uncertainty*, is

$$C_1 = E\left\{k\sigma_{u_i}\sigma_{u_j}\rho_{u_iu_j} \mid D_T\right\}.$$

This component can be interpreted as the uncertainty arising from the, by assumption, i.i.d. unexpected returns. It contributes the same per-period expected covariance at all investment horizons. To see this, notice that C_1 can be written more compactly as $kE\left\{\sigma_{u_iu_j} \mid D_T\right\}$. This also shows that C_1 contributes positively to the long-horizon predictive covariance when $i = j$ (and covariance is variance), but not necessarily so if $i \neq j$.

2.6.2 Mean reversion

The second component of Equation (16), labeled *mean reversion*, is

$$\begin{aligned} C_2 = &\underbrace{E\left\{k\sigma_{u_i}\sigma_{u_j}\left[b_i\bar{e}_i\rho_{v_iu_j}A_{\gamma_i}(k) + b_j\bar{e}_j\rho_{u_iv_j}A_{\gamma_j}(k)\right] \mid D_T\right\}}_{\textit{mean reversion about } x_s} \\ &+ \underbrace{E\left\{k\sigma_{u_i}\sigma_{u_j}\left[\bar{d}_i\rho_{\eta_iu_j}A_{\delta_i}(k) + \bar{d}_j\rho_{u_i\eta_j}A_{\delta_j}(k)\right] \mid D_T\right\}}_{\textit{mean reversion about } \pi_s}. \end{aligned}$$

This component reflects mean reversion in returns. In view of the specification of our predictive system in Section 2.4, we label the two parts of C_2 *mean reversion about* x_s and *mean reversion about* π_s, respectively.

2.6.3 Future μ_s uncertainty

The third component of Equation (16), labeled *future μ_s uncertainty*, is

$$C_3 = \underbrace{E\left\{k\sigma_{u_i}\sigma_{u_j}b_ib_j\bar{e}_i\bar{e}_j\rho_{v_iv_j}B_{\gamma_i,\gamma_j}(k) \mid D_T\right\}}_{\textit{future } x_s \textit{ uncertainty}} + \underbrace{E\left\{k\sigma_{u_i}\sigma_{u_j}\bar{d}_i\bar{d}_j\rho_{\eta_i\eta_j}B_{\delta_i,\delta_j}(k) \mid D_T\right\}}_{\textit{future } \pi_s \textit{ uncertainty}}$$
$$+ \underbrace{E\left\{k\sigma_{u_i}\sigma_{u_j}\left[b_i\bar{e}_i\bar{d}_j\rho_{v_i\eta_j}B_{\gamma_i,\delta_j}(k) + b_j\bar{d}_i\bar{e}_j\rho_{\eta_iv_j}B_{\delta_i,\gamma_j}(k)\right] \mid D_T\right\}}_{\textit{future } x_s \textit{ and } \pi_s \textit{ uncertainty}}.$$

This component reflects uncertainty about future expected returns. We label the three parts of C_3 *future x_s uncertainty*, *future π_s uncertainty*, and *future x_s and π_s uncertainty.*

2.6.4 Current π_s uncertainty

The fourth component of Equation (16), labeled *current π_s uncertainty*, is

$$C_4 = E\left\{\frac{1-\delta_i^k}{1-\delta_i}\frac{1-\delta_j^k}{1-\delta_j}q_{ij,T} \mid D_T\right\}.$$

This component reflects uncertainty about current unobserved predictors.

2.6.5 Estimation risk

The fifth and final component of Equation (16), labeled *estimation risk*, is

$$C_5 = Cov\left\{kE_{r_i} + \frac{1-\gamma_i^k}{1-\gamma_i}b_{i,T} + \frac{1-\delta_i^k}{1-\delta_i}c_{i,T},\, kE_{r_j} + \frac{1-\gamma_j^k}{1-\gamma_j}b_{j,T} + \frac{1-\delta_j^k}{1-\delta_j}c_{j,T} \mid D_T\right\}.$$

This covariance component involves $E_{r_s} = a_s + b_s\theta_s/(1-\gamma_s)$ and $b_{s,T} = a_s + b_sx_{s,T} - E_{r_s}$. The unconditional mean return and the spread between the on D_T conditional and the unconditional mean return, respectively.

2.6.6 Discussion

Parameter uncertainty plays a role in all five components of Equation (16). The first four components, C_1 through C_4, are expectations of random quantities due to uncertainty about ϕ (the parameters governing the joint dynamics of returns and predictors). If the values of these parameters were known to the investor, the expectation operators could be removed from these components. The last component, C_5, is the covariance of quantities whose randomness is also due to parameter uncertainty. In the absence of such uncertainty, the fifth component is zero, which is why we similar to Pástor and Stambaugh (2009) assign it the interpretation of estimation risk.

In our empirical analysis, we calculate the k-period covariance ratio defined as

$$CR(k) = \frac{Cov\left(r_{i,T}^{k}, r_{j,T}^{k} \mid D_T\right)}{k\, Cov\left(r_{i,T}^{1}, r_{j,T}^{1} \mid D_T\right)}, \tag{17}$$

and evaluate the importance of each of the five components of Equation (16) that vary with the investment horizon k to determine the shape of the long-horizon covariance curve.

3 Empirical Analysis

This section examines the long-horizon predictive variance of an investment strategy whereby a US investor buys a local currency bond while borrowing at the US dollar deposit rate. We first show that the excess returns from the strategy can be decomposed into three components: bond excess returns in local currency, the real interest rate differential between home and foreign currency, and the bilateral real exchange rate return. We then discuss the framework used to estimate the long-horizon predictive variance of the strategy returns.

3.1 Data and Preliminary Statistics

We focus on a sample of major countries relative to the US that exhibit a good degree of homogeneity and have relatively liquid and developed bond markets, i.e., Australia, Canada, Germany/Euro area, Japan, New Zealand, Norway, Sweden, Switzerland, and the UK. The main source of our dataset is Global Financial Data and the sample ranges from January 1800 and June 2017. The starting date varies, depending on data availability, across countries from January 1800 for the UK and January 1934 for Canada.

For each country in our sample, the set of returns r_t includes the bond excess return in local currency ($r_{1,t}$), the short-term real interest rate differential relative to the US ($r_{2,t}$), and the real exchange rate return vis-á-vis the US dollar ($r_{3,t}$). We construct $r_{1,t}$ using the log return on the 10-year government bond total return index ($y_t^\star$) minus the log return on the total return bills index, which we use as a proxy for the short-term rate ($i_t^\star$). Monthly data on the total return US bills index (i_t) are only available from December 1835 trough Global Financial data. We construct US short-term rates between January 1800 and December 1935 as in Siegel (1992) using yields from US commercial papers, US long-term bonds, UK long-term bonds, and UK short-term bonds. For the construction of $r_{2,t}$, we use the short-term rates defined above and the year-on-year log change on the consumer price index of the foreign country ($\rho_t^\star$) and the US (ρ_t), respectively. When monthly data are not available, we retrieve monthly observations by linearly interpolating lower-frequency data. For the calculation of the real exchange rate $r_{3,t}$, we combine the nominal exchange rate return (Δs_t) with the year-on-year log changes on the consumer price indices described above.

The set of observed predictors x_t includes the term spread, the long-term yield differential, and short-term yield differential. The term spread is measured as difference between the 10-year government bond yield minus the 3-month treasury bill yield. When the latter is not directly available, we use the log return on the total return bills index as a proxy. The long-term yield differential is quantified as the foreign country 10-year government bond yield minus the US 10-year government bond yield. The short-term yield differential is computed as the foreign country 3-month treasury bill yield minus the corresponding rate for the US.

Table 1 about here

We present the descriptive statics for both returns and observed predictors in Table 1. We also display descriptive statics for the cross-country average of r_t and x_t, respectively, which we label as "1/N" in the table. The bond excess return in local currency $rx_t^\star$ is generally positive across countries and ranges between 0.36 and 1.77 percent per annum for the UK and Canada, respectively. The exceptions are New Zealand and Switzerland for which the average bond excess return is negative and equal to -0.34 and -2.79 percent per annum, respectively. The standard deviation, reported in percentage per annum, is fairly large and goes from 8.35 for the UK to 4.99 for Switzerland. The short-term real interest rate differential rr_t is overall negative. For example, it is equal to -1.84 for Germany, -1.16 for the UK, and -0.16 for Japan. This evidence points out that the short-term real interest rate for the US has been, on average, higher than the short-term real interest rate abroad in our sample. The cross-sectional variation of rr_t as measured by the standard deviation is fairly small and we record the largest value of 2.26 percent per annum for Japan. The real exchange rate Δq_t evolves around zero, being negative for 5 countries and positive for the remaining 4 countries. We record the largest negative return for the UK (i.e., -1.16 percent per annum) and the largest positive value for Germany (i.e., 0.60 percent per annum). The standard deviation for Δq_t is generally larger than the standard deviation reported for $rx_t^\star$. For example, Australia displays a standard deviation of 16.38 percent per annum whereas New Zealand has a standard deviation of 11.72 percent per annum.

In addition to sample averages and standard deviations, we also measure the sample higher moments. The conventional measures of skewness and kurtosis, however, can be arbitrary large especially when the sample is contaminated by large values. This is the case when one works with long-span samples. Our sample is indeed are characterized by a number of events, including financial crisis, different monetary policy regimes as well as periods with fixed and floating exchange regimes. Rather than manually removing large values, we employ robust measures of skewness and kurtosis (e.g., Kim and White, 2004). In particular, the coefficient of skewness is defined as $skew = (\mu - Q_2)/\sigma$, where μ is the sample mean, Q_2 is the sample mode, and σ is the sample standard deviation. The centred coefficient of kurtosis, moreover, is computed as $kurt = ((F^{-1}_{0.975} + F^{-1}_{0.025})/(F^{-1}_{0.75} + F^{-1}_{0.25})) - 2.91$, where F^{-1} denotes the quantile of the empirical distribution. We find that the coefficient of skewness is by and large negative but small in size. In contrast, the coefficient of kurtosis is sizeable, especially for the real exchange rate return. We also compute the first-order serial correlation. We find that bond excess returns and real exchange rate returns have a relatively low coefficient of serial correlation whereas the real interest rate differentials display a very high coefficient.[4]

Table 1 also reports descriptive statistics of the observable predictors used to forecast bond excess returns, real interest rate differentials and real exchange returns. Term spreads are overall positive, denoting upward sloping yield curves for the majority of the countries. Nominal long- and short-term yield differentials are overall positive on average, suggesting that nominal long- and short-term nominal interest rates have been on average higher than the ones recorded in the US. As already reported in existing studies, all three predictors exhibit high persistence, with first-order serial correlation coefficients above 0.91.

4 Estimation and Results

In this section we describe our results pertaining to the estimation of the long-horizon predictive variance of the strategy returns as discussed in Section 3.2. After estimating for all countries in our sample the predictive system reported in Section 2.1, we compute the quantities discussed in Section 2.4 to obtain values of the predictive variance of monthly returns over an investment horizon up to 50 years. We also compute the results for a strategy that invests equally in all foreign bonds in our sample, which we label 1/N in the spirit of DeMiguel, Garlappi and Uppal (2009). Prior to discussing the main results regarding the predictive variance of the strategy returns, it is instructive to look at the estimates of the

[4]The implications of this statistical features are discussed further in the next Section 5.

predictive equation (8) reported in Table 2, in order to gauge the different degree of predictability exhibited by bond excess returns, interest rate differentials and real exchange rate returns.

Table 2 about here

The estimates of the ceteris paribus parameters b_1 associated with the predictability of bond excess returns in local currency suggest that in all cases bond excess returns are predictable by local-currency term spreads. In most of the countries the significance is at the 1 percent statistical level, corroborating the evidence already reported in the existing literature for the US Treasury bond market (e.g., Joslin et al., 2014). The recorded R^2 are also consistent with the ones recorded for the US at the monthly frequency (Gargano, Riddiough and Sarno, 2018). Similar evidence of predictability is found for the short-term real interest rate differentials, as all estimates b_2, with the exception of Japan, are statistically significant. Given the extreme persistence of the interest rate differential time series, the recorded R^2 are all large and exceeding 92 percent. The evidence of predictability of real exchange rate returns is less pervasive as, consistently with much empirical evidence, the estimates of the slope parameter b_3 are only significant in 4 cases out of 10 with only one case significant at the 1 percent level. Although there is compelling evidence of predictability for bond excess returns in local currency and interest rate differentials, the predictive variance of the strategy returns is likely to be affected by considerable uncertainty, as real exchange returns are not found predictable by short-term interest rate differentials and the degree of predictability of bond excess returns, measured by the R^2, is relatively small. We complete the analysis of the predictive model by reporting the model estimates for equations (9) and (10) of the main text in Tables 3 and 4, respectively.

Tables 3 and 4 about here

The evidence reported in Table 3 confirms the high persistence of all observable predictors, for all countries. In fact, the coefficients γ_i are all statistically significant at the 1 percent level and their value is close or exceed 0.94 across countries and predictors. Interestingly, the high persistence of observable predictors is mimicked by the persistence of the estimated unobserved predictors. In fact, although the coefficients δ_i reported in Table 4 are slightly smaller than the ones reported in Table 3, their value is still substantially large. This suggests that the predictable, potentially slow-moving, part of bond excess returns, real interest rate differentials and real exchange rates require multiple persistent predictors to be captured. The observable predictors employed in the empirical investigation, by themselves, are not able to capture the features of future interest rates and returns. Given the long sample period investigated, it is reasonable to hypothesize that those unobserved predictors may be associated with abrupt by persistent changes in monetary and exchange rate regimes experienced by the various countries during the past two decades.

Using the model estimates, we compute the predictive variances for all countries over the full sample and then report the predictive variance ratios in Figure 1.

FIGURE 1 ABOUT HERE

The long-term risk profile of investments in long-term local currency bonds exhibits in all cases an upward pattern: the longer the investment horizon, the larger the predictive variance. Over a 50-year horizon, the monthly predictive variance of the strategy returns increases by 200 to 300 percent across countries. For example, in the case of US investment in UK long-term bonds, the long-term predictive variance at 50 year horizon reaches the value of 4.24 percent per annum, a value that is comparable the long-horizon predictive variance of a US equity buy and hold strategy computed over the same sample period. This result also echoes early findings by Campbell and Viceira (2001, 2002) that show that long-term bonds are not very different from equities over longer investment horizons.

In order to understand the main driver of the long-horizon predictive variance of the strategy returns, we can we compute its constituents, in the spirit of the decomposition discussed in Sections 2.4 and 3.2. For the sake of exposition, we focus our discussion on the case of the 1/N investment strategy in order to gauge patterns that are common across countries over the sample period. However, with very few exceptions, the results are broadly comparable with the ones exhibited by investments in individual countries.

Figure 2 about here

The various constituents of the predictive variance are reported in the two panels of Figure 2. The top panel reports the constituents in terms of individual predictive variances of and pairwise covariances among the three components of the strategy returns. In light of this metric of assessment, the upward pattern exhibited by predictive variance of the strategy is mostly due to the individual predictive variances of the real exchange rate returns and the short-term interest rate differential, with the former playing a more prominent role. The predictive variance of bond excess returns in local currency does not play a substantial role, as its impact is relatively constant across investment horizons. Among the three predictive covariances, only the one between real interest rate differentials and real exchange rate returns is large and negative in sign. This finding suggests that predictive co-movement between bond excess returns in foreign currency and interest rates and exchange rate returns are less important in determining the long-term risk profile of the strategy. The results also suggest that the predictive co-movements between interest rates and real exchange rate are important in the long-run, as they tend to reduce the overall expected risk of the strategy, especially at longer horizons. This negative sign is also in line with the recent evidence reported in Engel (2016), whereby the correlation between long-term expected risk premia and real interest rate differential is negative.

The bottom panel of Figure 2 reports an alternative decomposition of the predictive variance that focuses on the mean reversion in returns and the set of uncertainties discussed in Section 2.4. The patterns exhibited by this decomposition are rather striking and they unambiguously assign a dominant, if not exclusive role, to the uncertainty about future returns. This type of uncertainty is found to be important for long horizons in the context of US equity markets (e.g., Pástor and Stambaugh, 2012). However, its impact on the predictive variance of returns is generally offset by the effect of mean reversion in returns, especially at short to medium horizons. In our context, the overall effect of the mean reversion components computed for each predictive variance and covariance in equation (12) is very close to zero for all investment horizons and, if anything, it exhibits a positive (rather than negative) value, adding to the long-horizon risk profile of the strategy.

Figure 3 about here

This last finding raises the question of why the mean reversion in returns, even in the presence of predictability of both bond excess returns and interest rate differentials, does not reduce the overall risk profile of the strategy at longer horizon. Further light can be shed on this issue by computing the mean reversion components of the predictive variance due to both the observable and unobserved predictors, plotted in Figure 3. The rather flat profile of the overall mean reversion component is due to the compensation that occurs between the mean reversions originating from the observable and unobserved predictors. While unobserved predictors exhibit a mean reversion effect on the predictive variance of the strategy that is progressively more negative, albeit small in value, as the investment horizon increases, observable predictors generate mean reversion effects that increase over the investment horizon. The sum of the two effects is therefore slightly positive, as the mean reversion of observable predictors is larger in absolute value, but rather flat at medium to long investment horizons. The lack of negative mean reversion effects on the predictive variance is at odd with the results reported in the existing literature (Pástor and Stambaugh, 2012; Avramov et al., 2018). However it can be easily rationalized on the ground that in a multi-asset context, like the one studied in this paper, the mean reversion effects originating from predictive covariances ought not be necessarily negative.

Figure 4 about here

In fact, if we plot the mean reversion components computed from all predictive variances and covariances needed to compute the predictive variance of the overall strategy as in Section 3.2, it is evident that the mean reversion components for bond excess returns and interest rate differentials, for which we have found evidence of predictability, are indeed negative. However, these effects are more than offset by the mean reversion components associated with the various predictive covariances, especially the one between real interest rate differentials and bond excess returns in local currency, that are large and positive in value. It is important to emphasize that in a multi-asset context, and in the presence of predictive covariances, there is no certainty that any evidence of predictability leads to mean reversion effects reducing the predictive variance of the overall strategy. In fact, it all depends on the potential but not certain occurrence of mean reverting effects to expected returns in one asset market which are due to shocks originating in other asset markets. Exploring further, we can understand the sources of the uncertainty about future expected returns by performing a set of decompositions similar to the ones reported in Figures 3 and 4.

FIGURE 5 ABOUT HERE

Figure 5 shows the uncertainty about future expected returns components of the predictive variance associated with both observable and unobserved predictors, and their joint interaction. The overall effect is again dominated by th0e component associated with unobserved predictors. In fact, albeit both increasing with the investment horizon, in the long run the impact on the predictive variance from the unobserved predictors is 3 times larger the one exhibited by observable predictors. Their joint interaction plays a minor role and it is only moderately negative for very long horizon.

FIGURE 6 ABOUT HERE

Figure 6 allows us to attribute the pattern shown in Figure 5 to uncertainty about future returns components originating from the predictive variances and covariances as in Section 3.2. It is evident that the dynamic range exhibited by the uncertainties associated with unobserved predictors is much larger than the one associated with observable ones. Nonetheless, in both cases, the major contributors are due to the uncertainty about future expected returns due to real interest rate differentials and real exchange rates returns. If we think of unobserved predictors as variables associated with changes in monetary and exchange rate regimes that have occurred over the past two centuries and are not already captured by the set of observable predictors, the evidence in Figure 6 suggests that those changes, especially their logical unpredictable effect on future returns, exert first order effects on the overall risk profile of the strategy. More specifically, given that the magnitude of the effects for both components associated with real interest rate differentials and real exchange rate returns is fairly comparable at longer horizons, both changes in monetary and exchange rate regimes play an important role when investing in local currency bonds in the long run.

5 Conclusions

This study investigates the long-horizon risk profile of an international bond strategy whereby a US investor earns excess returns from an unhedged position in a portfolio of long-term foreign bonds funded at the domestic risk-free rate. After showing that the excess returns from this strategy can be broken up into bond excess returns in local currency, the short-term real interest differential between the foreign and domestic currency and the bilateral real exchange rate return, we derive in closed-form and estimate an informative decomposition for the predictive variance of the strategy returns over different investment horizons ranging between 1 month and 50 years.

The empirical investigation is carried out by exploring investments in long-term bonds denominated in major currencies over the past two centuries. We compute the long-term predictive variances in an environment with imperfect predictability, by allowing unobserved predictors to join in a menu of observable predictors to help forecast the strategy returns at different horizons. We find a number of interesting results: First, over the full sample period and across all countries, the predictive variance of the bond investment strategy is found to be increasing with the investment horizon and this is mostly due to an upward sloping predictive variance of both short-term interest rate differentials and real exchange rate returns. Overall, the predictive variance of real exchange rate returns exhibits the largest long-horizon value, followed by interest rate differentials while the predictive variance of bond excess returns in foreign currency does not vary much across investment horizons. Predictive co-movement between bond excess returns in foreign currency and interest rates and exchange rate returns are less important in determining the long-term risk profile of the strategy but the predictive co-movements between interest rates and real exchange rate are important in the long-run, as they tend to reduce the overall expected risk of the strategy, especially at longer horizons.

Second, we observe that in all cases the uncertainty about future returns plays unambiguously the leading role in determining the shape of the predictive variance of the strategy returns at all investment horizons. All other components, especially the one associated with the mean reversion due to return predictability, are negligible in size and do not provide any improvement in the risk profile of the strategy. This suggests that, even when bond returns, interest rate and foreign exchange are found to be predictable, the range of uncertainties present in the predictive variance more than offset any potential benefit originating from any potential mean reversion in returns. Furthermore, in the multi asset context studied in this paper, there is no certainty that any evidence of predictability leads to mean reversion effects reducing the predictive variance of the overall strategy. In fact, it all depends on the potential but not certain occurrence of mean reverting effects to expected returns in one asset market which are due to shocks originating in other asset markets.

Across all currency pairs investigated, the uncertainty about future returns is mostly due to the component pertaining to the unobserved predictors. The uncertainty associated with the expected future values of the observable predictors is non-negligible but substantially smaller than the one recorded for unobserved predictors. Furthermore, as the uncertainty about future returns originating from interest rate differentials and exchange rate returns is fairly similar, the shape of the predictive variance over longer horizons can be interpreted as spurring from changes in monetary and exchange rate regimes that are not captured by the set of observable predictors.

We cannot detect tangible effects of mean reversion on the long-horizon predictive variance as, for most of the currencies considered, the expected negative impact of the mean reversion component on the predictive variance (only detected for unobserved predictors) is completely offset by a similar but positive impact of the mean reversion component originating from observable predictors. As the overall effect cancels out, the shape of strategy's the predictive variance is essentially due to the interplay between the various sources of uncertainties and therefore increases with the investment horizon.

References

Acharya, Viral V. and Sascha Steffen, "The "Greatest" Carry Trade Ever? Understanding Eurozone Bank Risks," *Journal of Financial Economics*, 2015, *115*, 215–236.

Allen, Kate, "Austria Sells Record Largest €3.5bn Century Bond," *Financial Times*, 2017, *https://www.ft.com/content/1d4f47e2-97a7-11e7-b83c-9588e51488a0.*

Andrews, Spencer, Riccardo Colacito, Mariano Massimiliano Croce, and Federico Gavazzoni, "Concealed Carry," *Working Paper*, 2021, *University of North Carolina at Chapel Hill.*

Avramov, Doron, Scott Cederburg, and Katarína Lučivjanská, "Are Stocks Riskier over the Long Run? Taking Cues from Economic Theory," *Review of Financial Studies*, 2018, *31*, 556–594.

Barberis, Nicholas, "Investing for the Long Run when Returns Are Predictable," *Journal of Finance*, 2000, *55*, 225–264.

Campbell, John Y., "A Variance Decomposition for Stock Returns," *Economic Journal*, 1991, *101*, 157–179.

_ **and Luis M. Viceira**, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford: Oxford University Press, 2002.

_ **and** _ , "The Term Structure of the Risk-Return Tradeoff," *Financial Analysts Journal*, 2005, *61*, 34–44.

_ **and Luis. Viceira**, "Who Should Buy Long-Term Bonds?," *American Economic Review*, 2001, *91*, 99–127.

_ **, Yeung Lewis Chan, and Luis M. Viceira**, "A Multivariate Model of Strategic Asset Allocation," *Journal of Financial Economics*, 2003, *67*, 41–80.

Carter, C. K. and R. Kohn, "On Gibbs Sampling for State Space models," *Biometrika*, 1994, *81*, 541–553.

Cenedese, Gino, Pasquale Della Corte, and Tianyu Wang, "Currency Mispricing and Dealer Balance Sheets," *Journal of Finance*, 2021, *76*, 2763–2803.

Cochrane, John H. and Monika Piazzesi, "Bond Risk Premia," *American Economic Review*, 2005, *95*, 138–160.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, "Optimal Versus Naive Diversification: How Inefficient is the 1-N Portfolio Strategy?," *Review of Financial Studies*, 2009, *22*, 1915–1953.

Engel, Charles, "Exchange Rates, Interest Rates, and the Risk Premium," *American Economic Review*, 2016, *106*, 436–74.

_ **and Kenneth West**, "Exchange Rates and Fundamentals," *Journal of Political Economy*, 2005, *113*, 485–517.

Fama, Eugene and Robert R Bliss, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 1987, *77*, 680–92.

Gargano, Antonio, Steven J. Riddiough, and Lucio Sarno, "Volume and Excess Returns in Foreign Exchange," *Working*, 2018, *Working Paper.*

Greenwood, Robin, Samuel G Hanson, Jeremy C. Stein, and Adiagg Sunderam, "A Quantity-Driven Theory of Term Premiums and Exchange Rates," Working Paper, Harvard University 2019.

Hicks, John, *Value and Capital*, Oxford: Oxford University Press, 1946.

Ilmanen, Antti, "Time-Varying Expected Returns in International Bond Markets," *Journal of Finance*, 1995, *50*, 481–506.

IMF, "Global Financial Stability Report: Lower for Longer," *International Monetary Fund*, 2019, *October.*

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, 2014, *69*, 1197–1233.

Keynes, John M., *A Treatise on Money*, London: Harcourt, Brace, and Co., 1930.

Kim, Tae-Hwan and Halbert White, "On More Robust Estimation of Skewness and Kurtosis," *Finance Research Letters*, 2004, *1*, 56–73.

Ludvigson, Sydney C. and Serena Ng, "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 2009, *22*, 5027–5067.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan, "The Term Structure of Currency Carry Trade Risk Premia," *American Economic Review*, 2019, *109.*

Maggiori, Matteo, "Financial Intermediation, International Risk Sharing, and Reserve Currencies," *American Economic Review*, 2017, *107*, 3038–3071.

_ **, Brent Neiman, and Jesse Schreger**, "International Currencies and Capital Allocation," *Journal of Political Economy*, 2020, *128*, 2019–2066.

Meese, Richard and Kenneth Rogoff, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?," *Journal of International Economics*, 1983, *14*, 3–24.

Merton, Robert, "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *Review of Economics and Statistics*, 1969, *51*, 247–57.

Modigliani, Franco and Richard Sutch, "Innovations in Interest Rate Policy," *American Economic Review*, 1966, *56*, 178–197.

Pástor, Lubos and Robert F. Stambaugh, "Predictive Systems: Living with Imperfect Predictors," *Journal of Finance*, 2009, *64*, 1583–1628.

_ and _ , "Are Stocks Really Less Volatile in the Long Run?," *Journal of Finance*, 2012, *67*, 431–478.

Samuelson, Paul, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics*, 1969, *51*, 239–46.

Siegel, Jeremy J., "The Real Rate of Interest from 1800–1990: A Study of the U.S. and the U.K.," *Journal of Monetary Economics*, 1992, *29*, 227–252.

_ , *Stocks for the Long Run*, New York: McGraw Hill, 2008.

Stambaugh, Robert F., "Predictive regressions," *Journal of Financial Economics*, 1999, *54*, 375–421.

Viceira, Luis. M. and Kevin Wang, "Global Portfolio Diversification for Long-Horizon Investors," *Working Paper*, 2018, *Harvard University.*

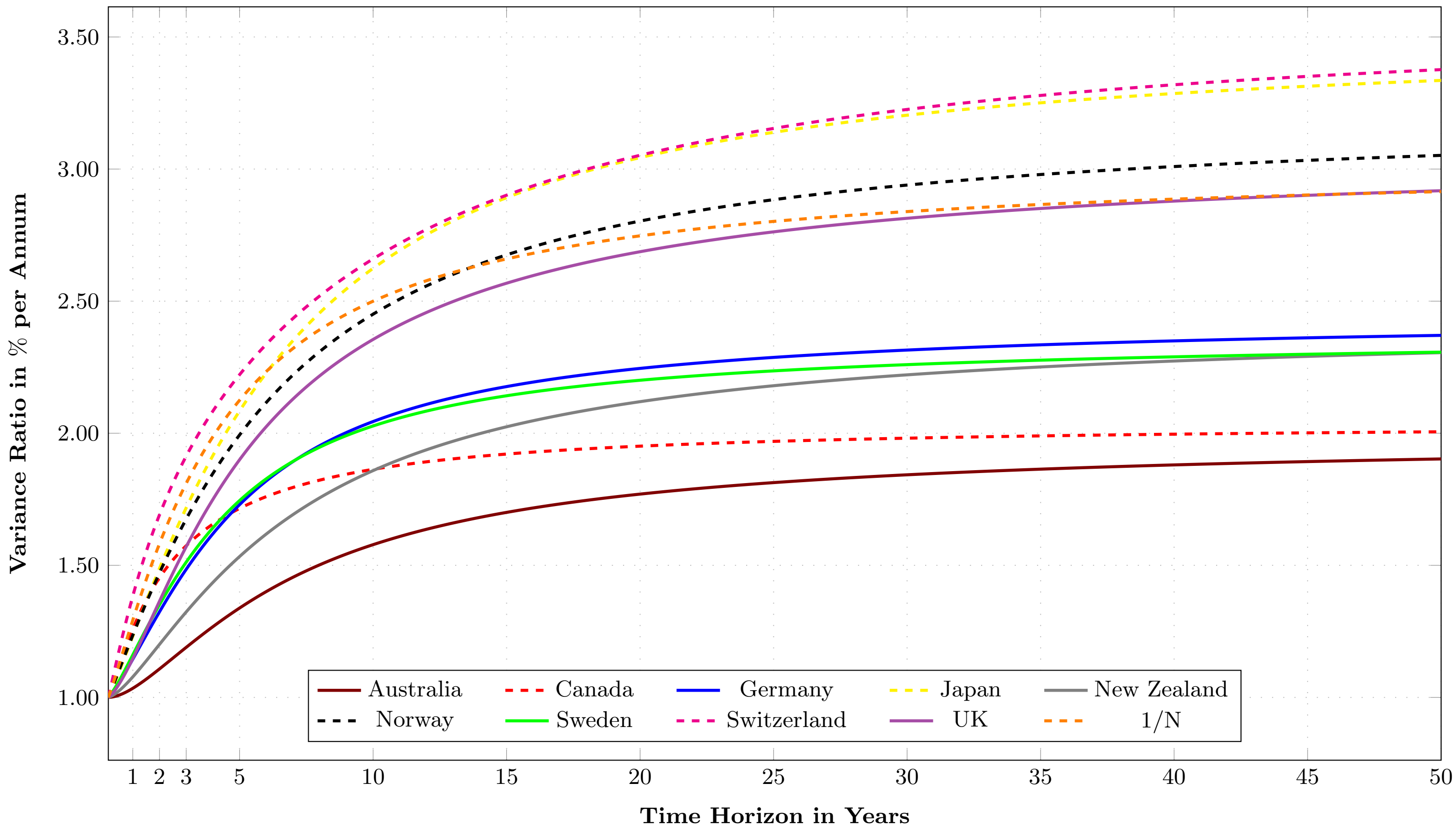

Figure 1: Predictive Variance Ratios

The figure shows the predictive variance ratio $Var\left(rx_T^k \mid D_T\right) / \left(kVar\left(rx_T^1 \mid D_T\right)\right)$ for local currency bond strategies, i.e., long-short strategies that invest in long-term foreign bonds while borrowing at the US short-term interest rate. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

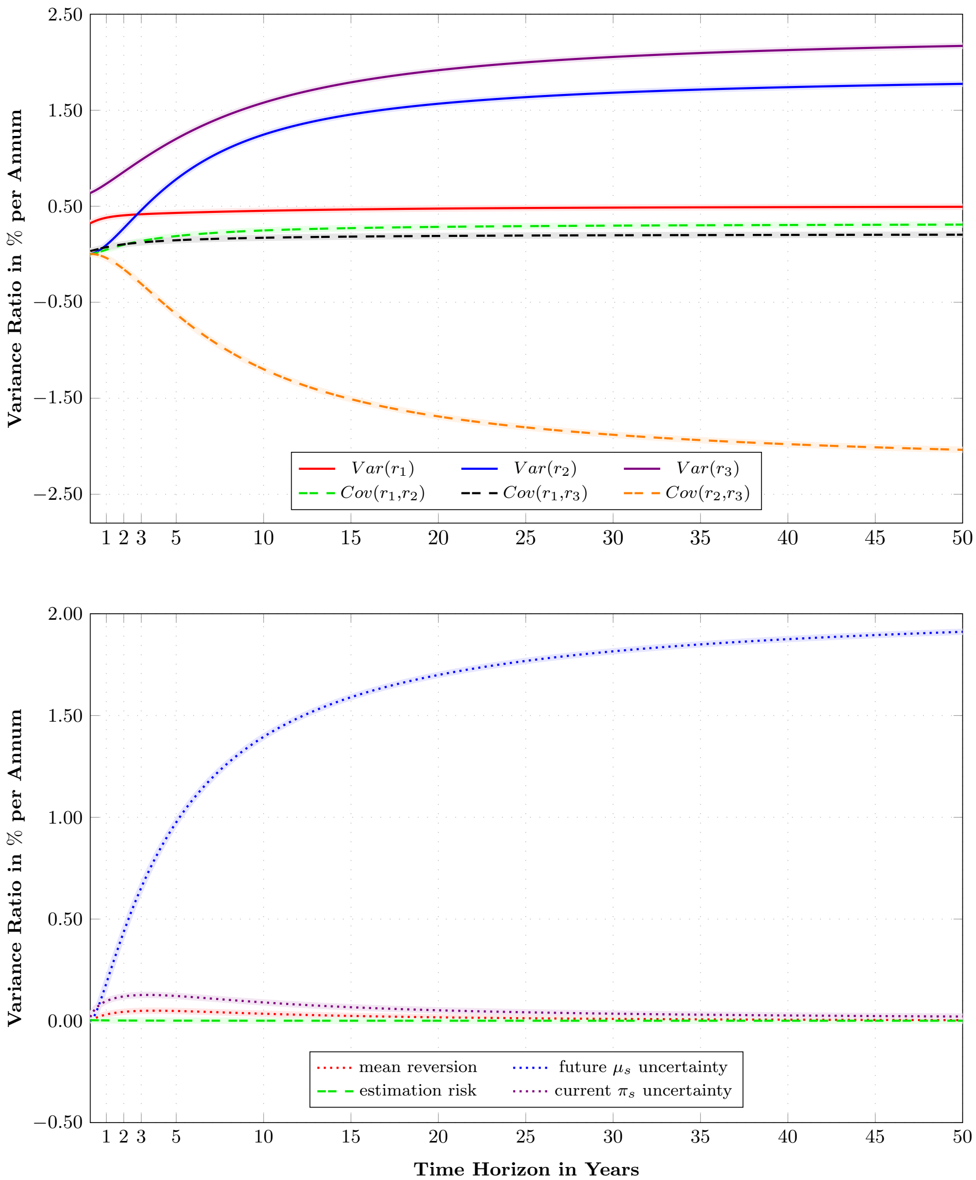

Figure 2: Decomposition of the Predictive Variance Ratio

The figure shows the decomposition of the predictive variance ratio $Var\left(rx_T^k \mid D_T\right) / \left(kVar\left(rx_T^1 \mid D_T\right)\right)$ for the 1/N international bond strategy. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

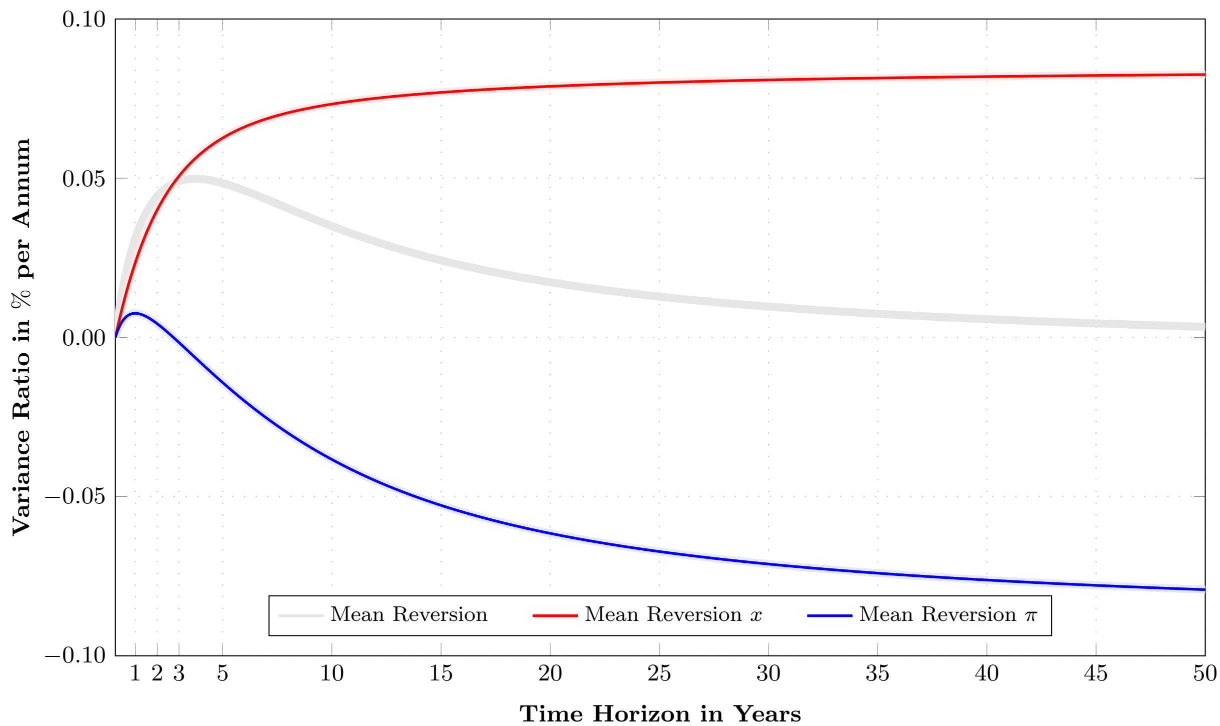

Figure 3: Mean Reversion Components

The figure shows the mean reversion components of the predictive variance ratio due to both observed and unobserved predictors for the 1/N international bond strategy. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

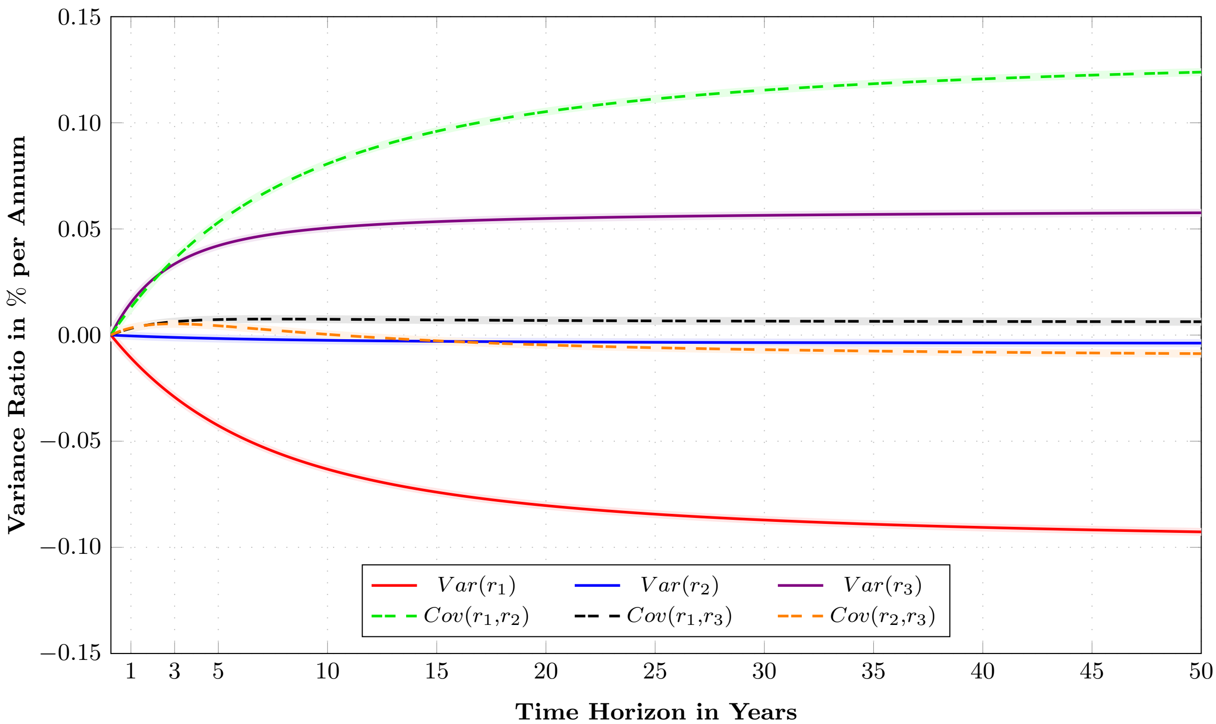

Figure 4: Mean Reversion of Observed Predictors

The figure shows the mean reversion components of the predictive variance ratio due to observed predictors for the 1/N international bond strategy. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

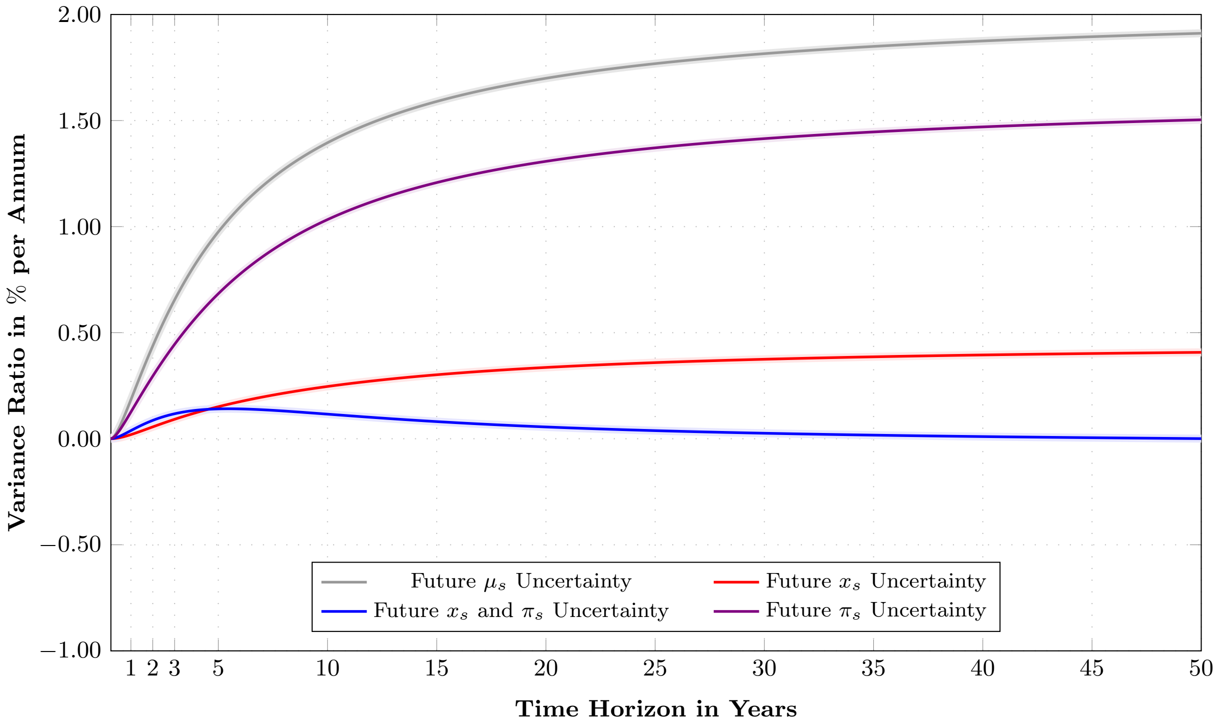

Figure 5: Future Uncertainty Components

The figure shows the future uncertainty components of the predictive variance ratio due to both observed and unobserved predictors for the 1/N international bond strategy. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

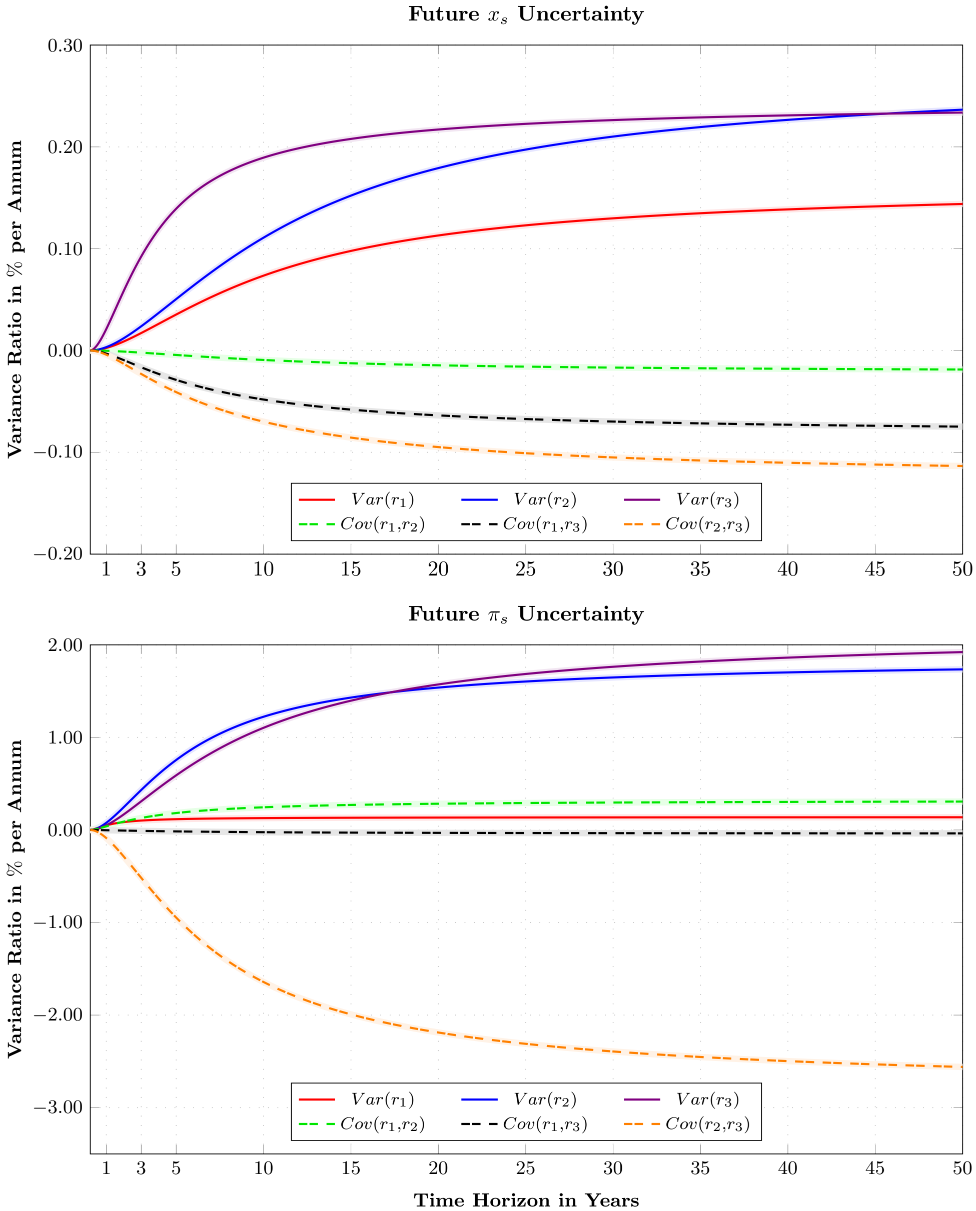

Figure 6: Decomposing Future Uncertainty Components

The figure decomposes the future uncertainty components of the predictive variance ratio due to both observed and unobserved predictors for the 1/N international bond strategy. The predictive variance is based on Equation (6) using the decomposition presented in Equation (16). The model parameters are reported in Tables 2–4 and are estimated via the Markov Chain Monte Carlo (MCMC) method discussed in Internet Appendix B. The sample consists of monthly data ranging from January 1800 and June 2017. Data are collected from Global Financial Data.

Table 1: Summary Statistics

This table presents descriptive statistics of dependent variables (bond excess returns, real interest rate differentials, and real exchange rate returns) and observed predictors (term spread, long-term yield differential, and short-term yield differential) for nine major countries relative to the US. 1/N denotes the naïve average across all countries. The mean and standard deviation (*sdev*) are in percentage per annum whereas skewness (*skew*) and excess kurtosis (*kurt*) are robust to the impact of outliers (e.g., Kim and White, 2004). ac_1 denotes the first-order autocorrelation coefficient. The sample period of monthly observations is reported in parenthesis. Data are sourced from Global Financial Data.

	mean	*sdev*	*skew*	*kurt*	ac_1	*mean*	*sdev*	*skew*	*kurt*	ac_1
	Australia (12/1862–06/2017)					**Canada** (01/1934–06/2017)				
Bond excess return	1.66	6.73	0.00	4.12	-0.11	1.77	5.79	-0.03	1.82	0.08
Real interest rate differential	-0.70	1.56	-0.08	1.18	0.98	0.55	0.74	0.06	1.60	0.95
Real exchange rate return	-0.60	16.38	-0.01	10.61	-0.27	-0.36	5.60	-0.03	3.52	-0.05
Term spread	1.39	0.33	-0.08	1.61	0.95	1.46	0.37	-0.11	0.60	0.96
Long-term yield differential	0.90	0.34	0.19	1.07	0.98	0.70	0.17	-0.17	0.55	0.95
Short-term yield differential	0.45	0.70	0.14	0.48	0.98	0.78	0.40	0.19	1.62	0.95
	Germany (12/1821–06/2017)					**Japan** (10/1882–06/2017)				
Bond excess return	1.43	6.99	-0.01	2.21	-0.01	1.13	7.76	0.00	3.28	-0.01
Real interest rate differential	-1.84	2.15	-0.18	2.93	0.97	-0.16	2.26	-0.01	2.89	0.97
Real exchange rate return	0.60	10.26	0.00	5.72	0.12	0.32	9.10	-0.03	4.47	0.09
Term spread	1.28	0.41	-0.08	0.02	0.91	0.61	0.55	-0.25	0.09	0.96
Long-term yield differential	0.31	0.50	-0.01	-0.35	0.99	0.81	0.82	-0.24	-0.38	0.99
Short-term yield differential	-0.29	0.65	-0.02	0.46	0.93	1.28	0.96	-0.21	-0.21	0.98
	New Zealand (01/1923–06/2017)					**Norway** (04/1822–06/2017)				
Bond excess return	-0.34	7.57	-0.01	3.08	0.09	0.58	8.15	0.04	2.30	0.03
Real interest rate differential	1.40	1.10	-0.20	-0.25	0.96	-0.76	1.87	-0.01	1.12	0.97
Real exchange rate return	0.07	11.72	-0.03	8.68	0.03	0.20	9.85	-0.01	5.60	0.11
Term spread	-0.50	0.61	-0.08	-0.68	0.95	0.32	0.49	0.07	-0.76	0.98
Long-term yield differential	1.16	0.49	0.07	4.08	0.97	0.44	0.35	0.03	0.44	0.98
Short-term yield differential	3.15	0.92	0.04	1.38	0.97	0.78	0.63	-0.01	0.73	0.95
	Sweden (10/1853–06/2017)					**Switzerland** (12/1899–06/2017)				
Bond excess return	0.88	5.30	0.01	2.76	0.03	-2.79	4.99	-0.06	2.21	0.11
Real interest rate differential	-0.71	1.63	-0.02	1.52	0.95	-0.27	1.04	0.01	1.19	0.96
Real exchange rate return	-0.23	8.26	-0.02	7.12	0.07	0.57	8.97	0.03	7.21	0.06
Term spread	0.98	0.31	0.08	1.81	0.93	1.00	0.33	-0.14	0.52	0.96
Long-term yield differential	0.60	0.29	-0.02	1.61	0.97	-0.78	0.64	-0.29	0.68	0.99
Short-term yield differential	0.46	0.57	0.20	1.03	0.96	-0.59	0.64	-0.22	0.69	0.98
	UK (01/1800–06/2017)					**1/N** (01/1800–06/2017)				
Bond excess return	0.36	8.35	0.00	2.36	-0.11	0.41	5.06	0.01	2.15	0.05
Real interest rate differential	-1.16	1.91	-0.10	1.72	0.97	-1.25	1.56	-0.09	1.10	0.97
Real exchange rate return	-0.39	8.44	0.00	4.52	0.10	0.10	7.19	-0.01	4.61	0.05
Term spread	0.49	0.44	0.11	0.36	0.95	0.43	0.30	-0.01	-0.56	0.98
Long-term yield differential	0.07	0.47	0.08	0.62	0.99	0.15	0.30	-0.08	-0.34	0.99
Short-term yield differential	0.15	0.56	0.03	2.08	0.93	0.30	0.42	-0.01	1.02	0.93

Table 2: Posterior Estimates: Return Equations

This table presents the Bayesian posterior means of the parameters underlying Equation (8). The numbers in parentheses indicate the numerical standard error (NSE). The superscripts *, **, and *** indicate that the 90%, 95%, and 99% highest posterior density (HPD) regions, respectively, do not contain zero. The posterior means are obtained via a Gibbs Sampling algorithm that runs for 100,000 iterations (after an initial burn-in of 100,000 iterations) and keeps one in ten iterations. The sample period for each country is presented in Table 1 and consists of monthly observations. Data are sourced from Global Financial Data.

	a_1	a_2	a_3	b_1	b_2	b_3	R_1^2	R_2^2	R_3^2
Australia	0.220	-1.138***	-0.938	1.119**	0.446**	0.665	2.2	96.3	0.5
	(0.013)	(0.010)	(0.006)	(0.015)	(0.011)	(0.016)			
Canada	-1.093	-0.277	-0.911	1.920***	1.190**	0.521	6.6	92.4	4.0
	(0.008)	(0.017)	(0.010)	(0.008)	(0.026)	(0.021)			
Germania	-0.648	-1.748***	0.883	1.601***	-0.455***	0.801	1.6	94.9	4.6
	(0.011)	(0.003)	(0.006)	(0.011)	(0.004)	(0.027)			
Japan	0.375	-0.263	0.024	1.201***	0.072	-0.135	1.7	95.2	6.9
	(0.006)	(0.005)	(0.007)	(0.009)	(0.003)	(0.014)			
New Zealand	0.114	0.387***	-3.921***	0.904*	0.841***	1.132***	3.1	93.5	2.3
	(0.007)	(0.005)	(0.008)	(0.021)	(0.003)	(0.009)			
Norway	0.079	0.244	-0.525	1.843***	-2.666***	0.983*	2.6	94.8	4.1
	(0.005)	(0.006)	(0.008)	(0.008)	(0.013)	(0.025)			
Sweden	-0.122	0.087	-0.357	0.974**	-1.538***	0.177	1.8	92.7	3.8
	(0.013)	(0.008)	(0.007)	(0.017)	(0.014)	(0.020)			
Switzerland	-4.729***	-0.077	0.865	1.956***	0.285***	0.434	15.3	93.3	3.3
	(0.0068)	(0.003)	(0.006)	(0.009)	(0.002)	(0.012)			
UK	-0.288	-1.016***	-0.495	1.267***	-2.425***	1.353**	0.8	94.5	3.3
	(0.007)	(0.004)	(0.006)	(0.018)	(0.009)	(0.024)			
1/N	-0.074	-0.980***	-0.228	1.147***	-1.671***	1.412**	3.2	94.0	4.1
	(0.004)	(0.003)	(0.006)	(0.009)	(0.009)	(0.023)			

Table 3: Posterior Estimates: Observed Predictor Equations

This table presents the Bayesian posterior means of the parameters underlying Equation (9). The numbers in parentheses indicate the numerical standard error (NSE). The superscripts *, **, and *** indicate that the 90%, 95%, and 99% highest posterior density (HPD) regions, respectively, do not contain zero. The posterior means are obtained via a Gibbs Sampling algorithm that runs for 100,000 iterations (after an initial burn-in of 100,000 iterations) and keeps one in ten iterations. The sample period for each country is presented in Table 1 and consists of monthly observations. Data are sourced from Global Financial Data.

	θ_1	θ_2	θ_3	γ_1	γ_2	γ_3	R_1^2	R_2^2	R_3^2
Australia	0.057***	0.026***	0.015***	0.958***	0.971***	0.971***	90.2	95.5	95.6
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Canada	0.079***	0.035***	0.031**	0.945***	0.948***	0.960***	91.8	89.3	90.1
	(0.002)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Germania	0.075***	0.002	-0.011	0.941***	0.987***	0.964***	83.0	98.5	86.6
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Japan	0.017	0.005	0.019	0.976***	0.988***	0.981***	91.2	98.6	96.5
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
New Zealand	-0.016	0.030**	0.097***	0.963***	0.975***	0.969***	90.5	94.9	94.2
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Norway	0.009	0.007	0.032**	0.975***	0.981***	0.956***	95.4	96.5	90.7
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Sweden	0.065***	0.018***	0.020	0.935***	0.970***	0.957***	86.8	94.5	91.3
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
Switzerland	0.042***	-0.009	-0.010	0.961***	0.988***	0.984***	92.3	98.8	96.0
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
UK	0.019*	0.001	0.006	0.962***	0.982***	0.958***	90.2	97.7	86.8
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			
1/N	0.010**	0.003	0.017	0.979***	0.981***	0.943***	95.7	97.1	85.5
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)			

Table 4: Posterior Estimates: Unobserved Predictor Equations

This table presents the Bayesian posterior means of the parameters underlying Equation (10). The numbers in parentheses indicate the numerical standard error (NSE). The superscripts *, **, and *** indicate that the 90%, 95%, and 99% highest posterior density (HPD) regions, respectively, do not contain zero. The posterior means are obtained via a Gibbs Sampling algorithm that runs for 100,000 iterations (after an initial burn-in of 100,000 iterations) and keeps one in ten iterations. The sample period for each country is presented in Table 1 and consists of monthly observations. Data are sourced from Global Financial Data.

	δ_1	δ_2	δ_3	R_1^2	R_2^2	R_3^2
Australia	0.951***	0.969***	0.935***	91.1	95.0	95.1
	(0.001)	(0.001)	(0.003)			
Canada	0.824***	0.949***	0.893***	67.4	93.9	89.8
	(0.003)	(0.001)	(0.002)			
Germania	0.820***	0.959***	0.956***	64.3	94.0	94.7
	(0.004)	(<.001)	(0.001)			
Japan	0.849***	0.965***	0.955***	82.0	94.7	95.9
	(0.005)	(<.001)	(0.001)			
New Zealand	0.918***	0.954***	0.929***	91.8	91.6	89.2
	(0.003)	(<.001)	(0.001)			
Norway	0.803***	0.966***	0.965***	58.5	95.0	96.0
	(0.004)	(<.001)	(<.001)			
Sweden	0.844***	0.959***	0.949***	76.5	94.3	93.0
	(0.004)	(<.001)	(0.001)			
Switzerland	0.869***	0.947***	0.946***	88.4	92.4	94.2
	(0.001)	(0.001)	(0.001)			
UK	0.883***	0.966***	0.973***	78.5	95.4	96.7
	(0.003)	(<.001)	(<.001)			
1/N	0.823***	0.963***	0.971***	70.4	94.2	94.9
	(0.002)	(<.001)	(<.001)			

Internet Appendix to

Currency Risk in the Long Run

(not for publication)

Abstract

We present supplementary results not included in the main body of the paper.

A Decomposition of $Cov\left(r_{i,T}^k, r_{j,T}^k \mid D_T\right)$

In Equation (16), we show that the long-horizon predictive covariance can be decomposed into five main components: *i.i.d. uncertainty*, *mean reversion*, *future* μ_s *uncertainty*, *current* π_s *uncertainty*, and *estimation risk*. Here we derive the expressions for these components in closed-form.

A.1 Derivation of $Cov\left(r_{i,T}^k, r_{j,T}^k \mid \pi_T, \phi, D_T\right)$

Since $x_{s,t}$ (the observable predictor of asset s at time t) in Equation (9) follows a first-order autoregression with $-1 < \gamma_s < 1$, we can rewrite $x_{s,t}$ as

$$x_{s,t} = \frac{1}{b_s}(E_{r_s} - a_s) + \sum_{l=0}^{\infty} \gamma_s^l v_{s,t-l}, \tag{A.1}$$

whenever $b_s \neq 0$. Similarly, since $0 < \delta_s < 1$, we can rewrite $\pi_{s,t}$ (the unobserved predictor of asset s at time t) in Equation (10) as

$$\pi_{s,t} = \sum_{l=0}^{\infty} \delta_s^l \eta_{s,t-l}. \tag{A.2}$$

From equations (8)–(11), the k-period return of asset s can be written as

$$\begin{aligned} r_{s,T+k} =& a_s + (1-\gamma_s^{k-1})(E_{r_s} - a_s) + b_s \gamma_s^{k-1} x_{s,T} + b_s \sum_{l=1}^{k-1} \gamma_s^{k-l-1} v_{s,T+l} \\ &+ \delta_s^{k-1} \pi_{s,T} + \sum_{l=1}^{k-1} \delta_s^{k-l-1} \eta_{s,T+l} + u_{s,T+k}. \end{aligned} \tag{A.3}$$

The k-period return from period $T+1$ through period $T+k$ is then

$$r_{s,T}^k = \sum_{l=1}^{k} r_{s,T+l} = kE_{r_s} + \frac{1-\gamma_s^k}{1-\gamma_s}(a_s + b_s x_{s,T} - E_{r_s}) + b_s \sum_{l=1}^{k-1} \frac{1-\gamma_s^{k-l}}{1-\gamma_s} v_{s,T+l}$$

$$+\frac{1-\delta_s^k}{1-\delta_s}\pi_{s,T}+\sum_{l=1}^{k-1}\frac{1-\delta_s^{k-l}}{1-\delta_s}\eta_{s,T+l}+\sum_{l=1}^{k}u_{s,T+l}. \tag{A.4}$$

The conditional covariance $Cov\left(r_{i,T}^k, r_{j,T}^k \mid \pi_T, \phi, D_T\right)$ of the k-period returns $r_{i,T}^k$ and $r_{j,T}^k$ can be obtained from (A.4) as

$$\begin{aligned}
&Cov\left(r_{i,T}^k, r_{j,T}^k \mid \pi_T, \phi, D_T\right) = k\sigma_{u_i u_j} + \\
&b_i b_j \frac{\sigma_{v_i v_j}}{(1-\gamma_i)(1-\gamma_j)}\left(k-1-\gamma_i\frac{1-\gamma_i^{k-1}}{1-\gamma_i}-\gamma_j\frac{1-\gamma_j^{k-1}}{1-\gamma_j}+\gamma_i\gamma_j\frac{1-\gamma_i^{k-1}\gamma_j^{k-1}}{1-\gamma_i\gamma_j}\right)+ \\
&b_i\frac{\sigma_{v_i\eta_j}}{(1-\gamma_i)(1-\delta_j)}\left(k-1-\gamma_i\frac{1-\gamma_i^{k-1}}{1-\gamma_i}-\delta_j\frac{1-\delta_j^{k-1}}{1-\delta_j}+\gamma_i\delta_j\frac{1-\gamma_i^{k-1}\delta_j^{k-1}}{1-\gamma_i\delta_j}\right)+ \\
&b_j\frac{\sigma_{\eta_i v_j}}{(1-\delta_i)(1-\gamma_j)}\left(k-1-\delta_i\frac{1-\delta_i^{k-1}}{1-\delta_i}-\gamma_j\frac{1-\gamma_j^{k-1}}{1-\gamma_j}+\delta_i\gamma_j\frac{1-\delta_i^{k-1}\gamma_j^{k-1}}{1-\delta_i\gamma_j}\right)+ \\
&\frac{\sigma_{\eta_i\eta_j}}{(1-\delta_i)(1-\delta_j)}\left(k-1-\delta_i\frac{1-\delta_i^{k-1}}{1-\delta_i}-\delta_j\frac{1-\delta_j^{k-1}}{1-\delta_j}+\delta_i\delta_j\frac{1-\delta_i^{k-1}\delta_j^{k-1}}{1-\delta_i\delta_j}\right)+ \\
&b_i\frac{\sigma_{v_i u_j}}{1-\gamma_i}\left(k-1-\gamma_i\frac{1-\gamma_i^{k-1}}{1-\gamma_i}\right)+b_j\frac{\sigma_{u_i v_j}}{1-\gamma_j}\left(k-1-\gamma_j\frac{1-\gamma_j^{k-1}}{1-\gamma_j}\right)+ \\
&\frac{\sigma_{\eta_i u_j}}{1-\delta_i}\left(k-1-\delta_i\frac{1-\delta_i^{k-1}}{1-\delta_i}\right)+\frac{\sigma_{u_i\eta_j}}{1-\delta_j}\left(k-1-\delta_j\frac{1-\delta_j^{k-1}}{1-\delta_j}\right).
\end{aligned} \tag{A.5}$$

We then break up the above conditional covariance into the three sources given by equations (13)–(15) and write is as in Equation (12), where $\bar{d}_s$ and $\bar{e}_s$ $(s = i, j)$ arise, respectively, from the relations

$$\sigma_{\eta_s}^2 = \sigma_{\pi_s}^2(1-\delta_s^2) = \sigma_{r_s}^2(1-\delta_s^2)R_s^2\frac{\sigma_{\pi_s}^2}{\sigma_{r_s}^2-\sigma_{u_s}^2} = \sigma_{u_s}^2(1-\delta_s^2)\frac{R_s^2}{1-R_s^2}\frac{\sigma_{\pi_s}^2}{\sigma_{r_s}^2-\sigma_{u_s}^2},$$

and

$$\sigma_{v_s}^2 = \sigma_{x_s}^2(1-\gamma_s^2) = \sigma_{r_s}^2(1-\gamma_s^2)R_s^2\frac{\sigma_{x_s}^2}{\sigma_{r_s}^2-\sigma_{u_s}^2} = \sigma_{u_s}^2(1-\gamma_s^2)\frac{R_s^2}{1-R_s^2}\frac{\sigma_{x_s}^2}{\sigma_{r_s}^2-\sigma_{u_s}^2},$$

which hold when returns are predictable.

A.2 Derivation of $Cov\left[E\left(r^k_{i,T} \mid \pi_T, \phi, D_T\right),\, E\left(r^k_{j,T} \mid \pi_T, \phi, D_T\right) \mid D_T\right]$

We finally derive a closed-form expression for the second term of the right-hand side of (7). For ease of exposition, let $E^k_{s,T} = E\left(r^k_{s,T} \mid \pi_T, \phi, D_T\right)$. The covariance of $E^k_{i,T}$ and $E^k_{j,T}$ given D_T can be decomposed as

$$\begin{aligned} Cov\left(E^k_{i,T}, E^k_{j,T} \mid D_T\right) &= E\left[Cov\left(E^k_{i,T},\, E^k_{j,T} \mid \phi, D_T\right) \mid D_T\right] \\ &\quad + Cov\left[E\left(E^k_{i,T} \mid \phi, D_T\right), E\left(E^k_{j,T} \mid \phi, D_T\right) \mid D_T\right]. \end{aligned} \tag{A.6}$$

By (A.4),

$$E^k_{s,T} = kE_{r_s} + \frac{1-\gamma^k_s}{1-\gamma_s}(a_s + b_s x_{s,T} - E_{r_s}) + \frac{1-\delta^k_s}{1-\delta_s}\pi_{s,T}.$$

It follows that

$$E\left(E^k_{s,T} \mid \phi, D_T\right) = kE_{r_s} + \frac{1-\gamma^k_s}{1-\gamma_s}(a_s + b_s x_{s,T} - E_{r_s}) + \frac{1-\delta^k_s}{1-\delta_s}c_{s,T}, \tag{A.7}$$

and

$$Cov\left(E^k_{i,T}, E^k_{j,T} \mid \phi, D_T\right) = \frac{1-\delta^k_i}{1-\delta_i}\frac{1-\delta^k_j}{1-\delta_j}q_{ij,T}. \tag{A.8}$$

Substituting equations (A.7) and (A.8) into Equation (A.6) and adding the on D_T conditional expectation of Equation (12) gives Equation (16).

B Bayesian Estimation

B.1 Predictive System in Compact Form

The predictive system in Equations (8)–(11) can equivalently be written as

$$y_t = X_{t-1}\beta + Z\pi_{t-1} + \varepsilon_t \tag{A.9}$$

$$\pi_t = \delta\pi_{t-1} + \eta_t, \tag{A.10}$$

with

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma'_{\eta\varepsilon} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{bmatrix} \right), \tag{A.11}$$

where $y_t = [r_t, x_t]'$ is a $2m \times 1$ vector that stacks together returns and observed predictors, X_t is a $2m \times 4m$ band matrix with $X_{i,t} = [1, x_{i,t}]$ and β is a $4m \times 1$ vector,

$$X_t = \begin{bmatrix} X_{1,t} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & X_{m,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & X_{1,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & X_{m,t} \end{bmatrix}, \quad \beta = \begin{bmatrix} \text{vec}\begin{pmatrix} a' \\ \text{diag}(b)' \end{pmatrix} \\ \\ \text{vec}\begin{pmatrix} \theta' \\ \text{diag}(\gamma)' \end{pmatrix} \end{bmatrix},$$

$Z = [I, 0]'$ is a $2m \times m$ matrix, and $\varepsilon_t = [u_t, v_t]'$ is a $2m \times 1$ vector that collects shocks to unexpected returns and observed predictors. In our empirical analysis, we set $m = 3$.

B.2 An Equivalent Representation with Orthogonal Shocks

We also impose, as a robustness exercise, a negative prior on the covariance between u_t and η_t as in Pástor and Stambaugh (2012). It is then convenient to rewrite the system in equations (A.9)–(A.11) with orthogonal shocks to y_t and π_t such that we can impose a given prior distribution on the parameters determining the covariance between u_t and η_t. To this end, we define the $m \times 1$ zero-mean random vector

$$\zeta_t \equiv \eta_t - \Sigma_{\eta\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t,$$

such that $\zeta_t \perp \varepsilon_t$ and $\Sigma_{\zeta\zeta} = \Sigma_{\eta\eta} - \Sigma_{\eta\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma'_{\eta\varepsilon}$. Substitution into Equation (A.10) gives

$$\pi_t = \delta\pi_{t-1} + \Sigma_{\eta\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t + \zeta_t.$$

We can then re-express the predictive system in equations (A.9)–(A.11) as

$$y_t = X_{t-1}\beta + Z\pi_{t-1} + \varepsilon_t \tag{A.12}$$

$$\pi_t = N_{t-1}\varphi + \zeta_t \tag{A.13}$$

with

$$\begin{bmatrix} \varepsilon_t \\ \zeta_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{\zeta\zeta} \end{bmatrix}\right), \tag{A.14}$$

where N_t is a $m \times m(2m+1)$ band matrix with $N_{i,t} = [\pi_{i,t}, \kappa'_{t+1}]$, $\kappa_t = \Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_t$ is a $2m \times 1$ vector of scaled shocks, and φ is a $m(2m+1) \times 1$ vector

$$N_t = \begin{bmatrix} N_{1,t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & N_{m,t} \end{bmatrix}, \quad \varphi = \text{ vec}\begin{pmatrix} \text{diag}(\delta)' \\ \Sigma'_{\eta\varepsilon} \end{pmatrix}.$$

B.3 Summary of the Algorithm

We estimate $\phi = [\beta, \varphi, \Sigma_{\varepsilon\varepsilon}, \Sigma_{\zeta\zeta}]$, the unknown parameters of the predictive system in Equations (A.12)–(A.14), using a Gibbs Sampling algorithm that conditions on $D_T = \{y_1, \ldots, y_T\}$ and $\pi = [\pi_1, \ldots, \pi_T]'$, where π is sampled in one block using the *forward filtering, backward sampling* approach of Carter and Kohn (1994). We draw 100,000 iterations (beyond a burn-in sample of 100,000 iterations) from the conditional posterior distributions and then keep a draw every 10 draws for a total of 10,000 draws.

The algorithm consists of the following steps:

1. Initialize $\beta, \Sigma_{\varepsilon\varepsilon}, \Sigma_{\zeta\zeta}$, and π_0 using random draws from the prior distributions,

2. Sample $\varphi \mid D_T, \pi, \phi_{-\varphi}$ from a conditional Normal posterior distribution,

3. Sample $\Sigma_{\zeta\zeta}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\zeta\zeta}}$ from a conditional Wishart posterior distribution,

4. Sample $\beta \mid D_T, \pi, \phi_{-\beta}$ from a conditional Normal posterior distribution,

5. Sample $\Sigma_{\varepsilon\varepsilon}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\varepsilon\varepsilon}}$ from a conditional Wishart posterior distribution,

6. Sample $\pi \mid \phi$ using Carter and Kohn (1994),

7. Go to step 2 and continue until you reach a total of 200,000 iterations.

We now describe the prior distributions, the posterior distributions and the algorithm of Carter and Kohn (1994).

B.4 Prior Distributions

We define the following prior distributions:

- A truncated multivariate Normal distribution for β defined as

$$\beta \sim \mathcal{N}\left(\underline{b}, \underline{B}\right) \qquad \text{and} \qquad p(\beta) \propto \exp\left[-\frac{1}{2}(\beta - \underline{b})'\underline{B}^{-1}(\beta - \underline{b})\right] \cdot \mathbb{1}_{\beta \in \mathbb{R}},$$

  where $\mathbb{1}$ is an indicator function and $\mathbb{R}$ represents the acceptance region for β. For a, $\text{diag}(b)$, and θ, we use the sample OLS estimates as prior means while choosing a prior variance of 0.5. We obtain the OLS estimates by estimating the predictive system without π_t. For γ, we use a prior mean of 0.90 and a prior variance of 0.02. The distribution is then truncated to satisfy the stationarity requirement $|\gamma_s| < 1$. The off-diagonal elements of $\underline{B}$ are set equal to zero.

- A truncated multivariate Normal distribution for φ defined as

$$\varphi \sim \mathcal{TN}\left(\underline{g}, \underline{G}\right) \qquad \text{and} \qquad p(\varphi) \propto \exp\left[-\frac{1}{2}(\varphi - \underline{g})'\underline{G}^{-1}(\varphi - \underline{g})\right] \cdot \mathbb{1}_{\varphi \in \mathbb{R}}.$$

The prior hyperparameters of δ are identical to those of γ. Akin to Pástor and Stambaugh (2012), we wish to be informative about $\sigma_{u_s\eta_s}$. We set the prior mean and standard deviation of $\sigma_{u_s\eta_s}$ such that $\rho_{u_s\eta_s}$ has a prior mean of -0.50 and a prior standard deviation of 0.16 (i.e., a three standard deviation confidence interval between -1 and 0). For all other elements of φ, we use a prior mean of zero and a prior standard deviation of 1. The distribution is then truncated such that $0 < \delta_s < 1$ and $-1 < \rho_{u_s\eta_s} < 0$. The off-diagonal elements of $\underline{G}$ are set equal to zero.

- A Wishart distribution for $\Sigma_{\varepsilon\varepsilon}^{-1}$ defined as

$$\Sigma_{\varepsilon\varepsilon}^{-1} \sim \mathcal{W}\left(\underline{S}_{\varepsilon}^{-1}, \underline{s}_{\varepsilon}\right) \qquad \text{and} \qquad p(\Sigma_{\varepsilon\varepsilon}^{-1}) \propto \left|\Sigma_{\varepsilon\varepsilon}^{-1}\right|^{\frac{(\underline{s}_{\varepsilon}-2m-1)}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\varepsilon\varepsilon}^{-1}\underline{S}_{\varepsilon}\right)\right],$$

where $\underline{S}_{\varepsilon}^{-1}$ is a positive definite scale matrix, $\underline{s}_{\varepsilon} > 2m-1$ is the degree of freedom, and $E\left(\Sigma_{\varepsilon\varepsilon}\right) = \underline{S}_{\varepsilon}^{-1}/(\underline{s}_{\varepsilon} - 2m - 1)$. We set $\underline{s}_{\varepsilon} = 2m+2$ and $\underline{S}_{\varepsilon}^{-1}$ equal to the sample OLS estimate of $\Sigma_{\varepsilon\varepsilon}$ while setting the off-diagonal elements equal to zero. We obtain the OLS estimates by estimating the predictive system without π_t.

- A Wishart distribution for $\Sigma_{\zeta\zeta}^{-1}$ defined as

$$\Sigma_{\zeta\zeta}^{-1} \sim \mathcal{W}\left(\underline{S}_{\zeta}^{-1}, \underline{s}_{\zeta}\right) \qquad \text{and} \qquad p(\Sigma_{\zeta\zeta}^{-1}) \propto \left|\Sigma_{\zeta\zeta}^{-1}\right|^{\frac{(\underline{s}_{\zeta}-m-1)}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\zeta\zeta}^{-1}\underline{S}_{\zeta}\right)\right],$$

where $\underline{S}_{\zeta}^{-1}$ is a positive definite scale matrix, $\underline{s}_{\zeta} > m-1$ is the degree of freedom, and $E\left(\Sigma_{\zeta\zeta}\right) = \underline{S}_{\zeta}^{-1}/(\underline{s}_{\zeta} - m - 1)$. We set $\underline{s}_{\zeta} = m+2$ and $\underline{S}_{\zeta}^{-1}$ equal to an identity matrix.

- A Normal distribution for π_0 defined as

$$\pi_0 \sim \mathcal{N}\left(b_0, Q_0\right) \qquad \text{and} \qquad p(\pi_0) \propto \exp\left[-\frac{1}{2}(\pi_0 - b_0)' Q_0^{-1}(\pi_0 - b_0)\right].$$

We set each element of b_0 equal to zero while Q_0 is an identity matrix.

B.5 Posterior Distributions

The joint posterior distribution is defined as

$$p(\phi, \pi \mid D_T) \propto p(D_T \mid \phi, \pi) \times p(\pi \mid \phi) \times p(\phi) \tag{A.15}$$

where the likelihood $p(D_T \mid \phi, \pi)$ is defined as

$$\begin{aligned} p(D_T \mid \phi, \pi) &= \prod_{t=1}^{T} p(y_t \mid \pi_t, \phi) \\ &\propto |\Sigma_{\varepsilon\varepsilon}|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{T} (y_t - X_{t-1}\beta - Z\pi_{t-1})' \Sigma_{\varepsilon\varepsilon}^{-1} (y_t - X_{t-1}\beta - Z\pi_{t-1})\right], \end{aligned}$$

the joint prior distribution $p(\phi)$ of the unknown parameters is simply

$$p(\phi) = p(\beta)\, p(\varphi)\, p\left(\Sigma_{\varepsilon\varepsilon}^{-1}\right) p\left(\Sigma_{\zeta\zeta}^{-1}\right),$$

and the prior distribution $p(\pi \mid \phi)$ of the state vector π is

$$\begin{aligned} p(\pi \mid \phi) &= \prod_{t=1}^{T} p(\pi_t \mid \pi_{t-1}, \phi) \\ &\propto |\Sigma_{\zeta\zeta}|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{T} (\pi_t - N_{t-1}\varphi)' \Sigma_{\zeta\zeta}^{-1} (\pi_t - N_{t-1}\varphi)\right]. \end{aligned}$$

While the joint posterior distribution does not take a convenient form, the conditional distributions are easy to derive. We now show the conditional posterior distributions.

B.5.1 Conditional Posterior of β

Start from the joint posterior distribution in Equation (A.15), set $\widetilde{y}_t = y_t - Z\pi_{t-1}$ and write the conditional posterior distribution of β as

$$p(\beta \mid D_T, \pi, \phi_{-\beta}) \propto p(D_T \mid \phi, \pi) \times p(\beta)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\begin{array}{c}\sum_t \left(\widetilde{y}_t - X_{t-1}\beta\right)' \Sigma_{\varepsilon\varepsilon}^{-1}\left(\widetilde{y}_t - X_{t-1}\beta\right)\\ +(\beta-\underline{b})'\underline{B}^{-1}(\beta-\underline{b})\end{array}\right]\right\}\cdot \mathbb{1}_{\beta\in\mathbb{R}}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\begin{array}{c}\sum_t \widetilde{y}_t'\Sigma_{\varepsilon\varepsilon}^{-1}\widetilde{y}_t - 2\sum_t \beta' X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}\widetilde{y}_t\\ +\sum_t \beta' X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}X_{t-1}\beta\\ +\beta'\underline{B}^{-1}\beta - 2\underline{b}'\underline{B}^{-1}\beta + \underline{b}'\underline{B}^{-1}\underline{b}\end{array}\right]\right\}\cdot \mathbb{1}_{\beta\in\mathbb{R}}.$$

Remove the terms that do not contain β, set $\overline{B} = (\underline{B}^{-1} + \sum_t X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}X_{t-1})^{-1}$ and $\bar{b} = \overline{B}(\underline{B}^{-1}\underline{b} + \sum_t X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}\widetilde{y}_t)$, and obtain

$$p\left(\beta \mid D_T, \pi, \phi_{-\beta}\right) \;\; \propto \exp\left\{-\tfrac{1}{2}\left[(\beta-\bar{b})'\overline{B}^{-1}(\beta-\bar{b})\right]\right\}\cdot \mathbb{1}_{\beta\in\mathbb{R}}.$$

It follows that β has a conditional (truncated) normal posterior distribution

$$\beta \mid D_T, \pi, \phi_{-\beta} \sim \mathcal{N}\left(\bar{b}, \overline{B}\right)\cdot \mathbb{1}_{\beta\in\mathbb{R}},$$

with

$$\overline{B} = \left(\underline{B}^{-1} + \sum\nolimits_t X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}X_{t-1}\right)^{-1}$$

and

$$\bar{b} = \overline{B}\left[\underline{B}^{-1}\underline{b} + \sum\nolimits_t X_{t-1}'\Sigma_{\varepsilon\varepsilon}^{-1}\left(y_t - Z\pi_{t-1}\right)\right].$$

B.5.2 Conditional Posterior of $\Sigma_{\varepsilon\varepsilon}^{-1}$

Start from the joint posterior distribution, set $\widetilde{y}_t = y_t - X_{t-1}\beta - Z\pi_{t-1}$ and write the conditional posterior distribution of $\Sigma_{\varepsilon\varepsilon}^{-1}$ as

$$p\left(\Sigma_{\varepsilon\varepsilon}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\varepsilon\varepsilon}}\right) \;\propto\; \left|\Sigma_{\varepsilon\varepsilon}\right|^{-\frac{T}{2}}\exp\left[-\frac{1}{2}\sum\nolimits_t \widetilde{y}_t'\Sigma_{\varepsilon\varepsilon}^{-1}\widetilde{y}_t\right]\times\left|\Sigma_{\varepsilon\varepsilon}^{-1}\right|^{\frac{(s_\varepsilon-2m-1)}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\varepsilon\varepsilon}^{-1}\underline{S}_\varepsilon\right)\right]$$

$$\propto\; \left|\Sigma_{\varepsilon\varepsilon}^{-1}\right|^{\frac{(T+s_\varepsilon-2m-1)}{2}}\exp\left[-\frac{1}{2}\sum\nolimits_t \widetilde{y}_t'\Sigma_{\varepsilon\varepsilon}^{-1}\widetilde{y}_t\right]\times\exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\varepsilon\varepsilon}^{-1}\underline{S}_\varepsilon\right)\right].$$

Using the properties of the trace operator,[5] we can rewrite the conditional posterior of $\Sigma_{\varepsilon\varepsilon}^{-1}$ as

$$\begin{aligned} p\left(\Sigma_{\varepsilon\varepsilon}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\varepsilon\varepsilon}}\right) &\propto \left|\Sigma_{\varepsilon\varepsilon}^{-1}\right|^{\frac{(T+\underline{s}_\varepsilon-2m-1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma_{\varepsilon\varepsilon}^{-1}\left(\sum\nolimits_t \widetilde{y}_t\widetilde{y}_t' + \underline{S}_\varepsilon\right)\right]\right\} \\ &\propto \left|\Sigma_{\varepsilon\varepsilon}^{-1}\right|^{\frac{(\overline{s}_\varepsilon-2m-1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\overline{S}_\varepsilon\Sigma_{\varepsilon\varepsilon}^{-1}\right]\right\}, \end{aligned}$$

where $\overline{S}_\varepsilon = \underline{S}_\varepsilon + \sum_t \widetilde{y}_t\widetilde{y}_t'$ and $\overline{s}_\varepsilon = T + \underline{s}_\varepsilon$. It follows that $\Sigma_{\varepsilon\varepsilon}^{-1}$ has a conditional posterior Wishart distribution as

$$\Sigma_{\varepsilon\varepsilon}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\varepsilon\varepsilon}} \sim \mathcal{W}\left(\overline{S}_\varepsilon^{-1}, \overline{s}_\varepsilon\right),$$

with

$$\begin{aligned} \overline{S}_\varepsilon &= \underline{S}_\varepsilon + \sum\nolimits_t (y_t - X_{t-1}\beta - Z\pi_{t-1})(y_t - X_{t-1}\beta - Z\pi_{t-1})' \\ \overline{s}_\varepsilon &= T + \underline{s}_\varepsilon. \end{aligned}$$

B.5.3 Conditional Posterior of φ

Start from the joint posterior distribution and write the conditional posterior of φ as

$$\begin{aligned} p(\varphi \mid D_T, \pi, \phi_{-\varphi}) &\propto \exp\left\{-\frac{1}{2}\left[\begin{array}{c} \sum_t (\pi_t - N_{t-1}\varphi)'\Sigma_{\zeta\zeta}^{-1}(\pi_t - N_{t-1}\varphi) \\ +(\varphi-\underline{g})'\underline{G}^{-1}(\varphi-\underline{g}) \end{array}\right]\right\} \cdot \mathbb{1}_{\varphi\in\mathbb{R}} \\ &\propto \exp\left\{-\frac{1}{2}\left[\begin{array}{c} \sum_t \pi_t'\Sigma_{\zeta\zeta}^{-1}\pi_t - 2\sum_t \pi_t'\Sigma_{\zeta\zeta}^{-1}N_{t-1}\varphi \\ +\sum_t \varphi' N_{t-1}'\Sigma_{\zeta\zeta}^{-1}N_{t-1}\varphi \\ +\varphi'\underline{G}^{-1}\varphi - 2\underline{g}'\underline{G}^{-1}\varphi + \underline{g}'\underline{G}^{-1}\underline{g} \end{array}\right]\right\} \cdot \mathbb{1}_{\varphi\in\mathbb{R}}. \end{aligned}$$

[5] Recall that i) the trace of a scalar is the scalar itself, i.e., $\mathrm{tr}(a) = a$; ii) the trace operator is invariant under cyclic permutation, i.e., $\mathrm{tr}(AB) = \mathrm{tr}(BA)$; and iii) the trace is linear mapping, i.e., $\mathrm{tr}(A+B) = \mathrm{tr}(A)+\mathrm{tr}(B)$.

Remove the terms that do not contain φ, set $\overline{G} = (\underline{G}^{-1} + \sum_t N'_{t-1}\Sigma_{\zeta\zeta}^{-1}N_{t-1})^{-1}$ and $\overline{g} = \overline{G}\left(\underline{G}^{-1}\underline{g} + \sum_t N'_{t-1}\Sigma_{\zeta\zeta}^{-1}\pi_t\right)$, and obtain

$$p\left(\varphi \mid D_T, \pi, \phi_{-\varphi}\right) \quad \propto \exp\left\{-\tfrac{1}{2}\left[(\varphi - \overline{g})'\overline{G}^{-1}(\varphi - \overline{g})\right]\right\} \cdot \mathbb{1}_{\varphi \in \mathbb{R}}.$$

It follows that φ has a conditional posterior normal distribution as

$$\varphi \mid D_T, \pi, \phi_{-\varphi} \sim \mathcal{N}\left(\overline{g}, \overline{G}\right) \cdot \mathbb{1}_{\varphi \in \mathbb{R}},$$

where

$$\begin{aligned}\overline{G} &= \left(\underline{G}^{-1} + \sum\nolimits_t N'_{t-1}\Sigma_{\zeta\zeta}^{-1}N_{t-1}\right)^{-1} \\ \overline{g} &= \overline{G}\left(\underline{G}^{-1}\underline{g} + \sum\nolimits_t N'_{t-1}\Sigma_{\zeta\zeta}^{-1}\pi_t\right).\end{aligned}$$

B.5.4 Conditional Posterior of $\Sigma_{\zeta\zeta}^{-1}$

Start from the joint posterior distribution, set $\widetilde{\pi}_t = \pi_t - N_{t-1}\varphi$ and write the conditional posterior of $\Sigma_{\zeta\zeta}^{-1}$ as

$$\begin{aligned}p\left(\Sigma_{\zeta\zeta}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\zeta\zeta}}\right) \quad &\propto \quad |\Sigma_{\zeta\zeta}|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum\nolimits_t \widetilde{\pi}'_t\Sigma_{\zeta\zeta}^{-1}\widetilde{\pi}_t\right] \times \left|\Sigma_{\zeta\zeta}^{-1}\right|^{\frac{(\underline{s}_\zeta - m - 1)}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\zeta\zeta}^{-1}\underline{S}_\zeta\right)\right] \\ &\propto \quad \left|\Sigma_{\zeta\zeta}^{-1}\right|^{\frac{(T + \underline{s}_\zeta - m - 1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma_{\zeta\zeta}^{-1}\left(\sum\nolimits_t \widetilde{\pi}_t\widetilde{\pi}'_t + \underline{S}_\zeta\right)\right]\right\} \\ &\propto \quad \left|\Sigma_{\zeta\zeta}^{-1}\right|^{-\frac{(\overline{s}_\zeta - m - 1)}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma_{\zeta\zeta}^{-1}\overline{S}_\zeta\right)\right].\end{aligned}$$

It follows that $\Sigma_{\zeta\zeta}^{-1}$ has a conditional posterior Wishart distribution as

$$\Sigma_{\zeta\zeta}^{-1} \mid D_T, \pi, \phi_{-\Sigma_{\zeta\zeta}} \sim \mathcal{W}\left(\overline{S}_\zeta^{-1}, \overline{s}_\zeta\right),$$

where

$$\overline{S}_\zeta = \underline{S}_\zeta + \sum\nolimits_t (\pi_t - N_{t-1}\varphi)(\pi_t - N_{t-1}\varphi)'$$
$$\overline{s}_\zeta = T + \underline{s}_\zeta.$$

B.6 Sampling $\pi \mid \phi$ using Carter and Kohn (1994)

We sample the vector π in one block from the full conditional posterior distribution $p(\pi \mid D_T, \phi)$ using the *forward filtering, backward sampling* method of Carter and Kohn (1994). We will omit ϕ throughout this section for simplicity.

B.6.1 Forward Filtering

Recall the state space system in equations (A.9)–(A.11),

$$y_t = X_{t-1}\beta + Z\pi_{t-1} + \varepsilon_t$$
$$\pi_t = \delta\pi_{t-1} + \eta_t,$$

with

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma'_{\eta\varepsilon} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{bmatrix}\right).$$

Let $D_t = \{y_t, D_{t-1}\}$ be the information set at time t. Forward filtering consists of the following steps:

a) Initial condition at time $t-1$

$$\pi_{t-1} \mid D_{t-1} \sim \mathcal{N}(b_{t-1}, Q_{t-1}).$$

b) Prior at time t

$$\pi_t \mid D_{t-1} \sim \mathcal{N}\left(a_t, P_t\right),$$

where

$$\begin{aligned} a_t &= E\left(\pi_t \mid D_{t-1}\right) = \delta E\left(\pi_{t-1} \mid D_{t-1}\right) = \delta b_{t-1} \\ P_t &= Var\left(\pi_t \mid D_{t-1}\right) = \delta Var\left(\pi_{t-1} \mid D_{t-1}\right)\delta' + \Sigma_{\eta\eta} = \delta Q_{t-1}\delta' + \Sigma_{\eta\eta}. \end{aligned}$$

c) Prediction at time t

$$y_t \mid D_{t-1} \sim \mathcal{N}\left(f_t, S_t\right),$$

where

$$\begin{aligned} f_t &= E\left(y_t \mid D_{t-1}\right) = X_{t-1}\beta + ZE\left(\pi_{t-1} \mid D_{t-1}\right) = X_{t-1}\beta + Zb_{t-1} \\ S_t &= Var\left(y_t \mid D_{t-1}\right) = ZVar\left(\pi_{t-1} \mid D_{t-1}\right)Z' + \Sigma_{\varepsilon\varepsilon} = ZQ_{t-1}Z' + \Sigma_{\varepsilon\varepsilon}. \end{aligned}$$

d) Joint distribution at time t

$$\begin{bmatrix} y_t & \\ & \mid D_{t-1} \\ \pi_t & \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} f_t \\ a_t \end{bmatrix}, \begin{bmatrix} S_t & G_t \\ G_t' & P_t \end{bmatrix}\right),$$

where

$$\begin{aligned} G_t &= Cov\left(y_t, \pi_t \mid D_{t-1}\right) \\ &= ZVar\left(\pi_{t-1} \mid D_{t-1}\right)\delta' + Cov\left(\varepsilon_t, \eta_t \mid D_{t-1}\right) \\ &= ZQ_{t-1}\delta' + \Sigma_{\eta\varepsilon}'. \end{aligned}$$

e) Posterior at time t

$$\pi_t \mid D_t \sim \mathcal{N}\left(b_t, Q_t\right),$$

where

$$b_t = E\left(\pi_t \mid y_t, D_{t-1}\right) = a_t + G_t' S_t^{-1}\left(y_t - f_t\right)$$
$$Q_t = Var\left(\pi_t \mid y_t, D_{t-1}\right) = P_t - G_t' S_t^{-1} G_t.$$

The posterior hyperparameters b_t and Q_t are easily derived since $\pi_t \mid D_t$ is equivalent to $\pi_t \mid y_t, D_{t-1}$. Using the joint distribution of y_t and π_t presented above, it is easy to obtain the conditional distributions from a multivariate normal distribution.

B.6.2 Backward Filtering

The backward sampling method builds on the following Markov property

$$p\left(\xi_1, \ldots, \xi_T \mid D_T\right) = p\left(\xi_T \mid D_T\right) p\left(\xi_{T-1} \mid \xi_T, D_{T-1}\right) \times \cdots \times p\left(\xi_1 \mid \xi_2, D_1\right).$$

We sample π_T from $p\left(\pi_T \mid D_T\right)$ and then π_t from the conditional density $p\left(\xi_t \mid \xi_{t+1}, D_t\right)$ for $t = T-1, \ldots, 1$, where $\xi_t = [y_t, \pi_t]'$. We derive the conditional density $p\left(\xi_t \mid \xi_{t+1}, D_t\right)$ as follows

$$\underbrace{\begin{bmatrix} y_t \\ \pi_t \end{bmatrix}}_{\xi_t} = \underbrace{\begin{bmatrix} 0 & Z \\ 0 & \delta \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix}}_{\xi_{t-1}} + \underbrace{\begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 \\ X_{t-1} \end{bmatrix}}_{\Lambda_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}}_{e_t}$$

and recall that

$$\xi_{t+1} \mid D_t \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}}_{\overline{a}_{t+1}}, \underbrace{\begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}' & P_{t+1} \end{bmatrix}}_{\overline{A}_{t+1}} \right)$$

and

$$\xi_t \mid D_t \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} y_t \\ b_t \end{bmatrix}}_{\bar{b}_t}, \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & Q_t \end{bmatrix}}_{\overline{B}_t}\right).$$

The **joint density** is then straightforward to derive as

$$\begin{bmatrix} \xi_t & \\ & \mid D_t \\ \xi_{t+1} & \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{b}_t \\ \overline{a}_{t+1} \end{bmatrix}, \begin{bmatrix} \overline{B}_t & \overline{G}_t \\ \overline{G}_t' & \overline{A}_{t+1} \end{bmatrix}\right),$$

where

$$\begin{aligned} \overline{G}_t &= Cov\left(\xi_t, \xi_{t+1} \mid D_t\right) \\ &= Cov\left(\xi_t, M\xi_t + L\Lambda_t + e_{t+1} \mid D_t\right) \\ &= Var\left(\xi_t \mid D_t\right) M' \\ &= \overline{B}_t M'. \end{aligned}$$

Hence, we have obtained

$$\xi_t \mid \xi_{t+1}, D_t \sim \mathcal{N}\left(h_t, H_t\right),$$

where

$$\begin{aligned} h_t &= E\left(\xi_t \mid \xi_{t+1}, D_t\right) = \bar{b}_t + \overline{G}_t \overline{A}_{t+1}^{-1}\left(\xi_{t+1} - \overline{a}_{t+1}\right) \\ H_t &= Var\left(\xi_t \mid \xi_{t+1}, D_t\right) = \overline{B}_t - \overline{G}_t \overline{A}_{t+1}^{-1} \overline{G}_t'. \end{aligned}$$

B.6.3 Summary

We now summarize the forward filtering and backward filtering algorithm as

a) Prediction Equations

$$\begin{aligned} E\left(\pi_t \mid D_{t-1}\right) &: a_t = \delta b_{t-1} \\ Var\left(\pi_t \mid D_{t-1}\right) &: P_t = \delta Q'_{t-1}\delta + \Sigma_{\eta\eta} \\ E\left(y_t \mid D_{t-1}\right) &: f_t = X_{t-1}\beta + Zb_{t-1} \\ Var\left(y_t \mid D_{t-1}\right) &: S_t = ZQ_{t-1}Z' + \Sigma_{\varepsilon\varepsilon} \\ Cov\left(y_t, \pi_t \mid D_{t-1}\right) &: G_t = ZQ_{t-1}\delta + \Sigma'_{\eta\varepsilon}. \end{aligned}$$

b) Updating Equations

$$\begin{aligned} E\left(\pi_t \mid y_t, D_{t-1}\right) &: b_t = a_t + G'_t S_t^{-1}\left(y_t - f_t\right) \\ Var\left(\pi_t \mid y_t, D_{t-1}\right) &: Q_t = P_t - G'_t S_t^{-1} G_t. \end{aligned}$$

c) Sample π_T^* from $p\left(\pi_T \mid D_T, \phi\right)$

$$\pi_T^* \sim \mathcal{N}\left(b_T, Q_T\right).$$

d) Sample π_t^* from $p\left(\pi_t \mid \pi_{t+1}, D_t, \phi\right)$ starting from $t = T-1, \ldots, 1$

$$\pi_t^* \sim \mathcal{N}\left(h_{t,\pi}, H_{t,\pi}\right),$$

where

$$\begin{aligned} E\left(\xi_t \mid \xi_{t+1}, D_t\right) &: h_t = \overline{b}_t + \overline{G}_t \overline{A}_{t+1}^{-1}\left(\xi_{t+1} - \overline{a}_{t+1}\right) \\ Var\left(\xi_t \mid \xi_{t+1}, D_t\right) &: H_t = \overline{B}_t - \overline{G}_t \overline{A}_{t+1}^{-1} \overline{G}'_t, \end{aligned}$$

with

$$\overline{b}_t = \begin{bmatrix} y_t \\ b_t \end{bmatrix}, \quad \overline{G}_t = \overline{B}_t M', \quad \overline{B}_t = \begin{bmatrix} 0 & 0 \\ 0 & Q_t \end{bmatrix}, \quad M = \begin{bmatrix} 0 & Z \\ 0 & \delta \end{bmatrix},$$

$$\overline{A}_{t+1} = \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix}, \quad \overline{a}_{t+1} = \begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}, \quad \xi_{t+1} = \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \end{bmatrix}.$$

The generated vector $[\pi_1^*, \dots, \pi_T^*]'$ corresponds to a random draw from $p(\pi_1, \dots, \pi_T \mid D_T)$.