Not-for-Publication Appendix to "An International Dynamic Term Structure Model with Economic Restrictions and Unspanned Risks"

Gregory H. Bauer
Bank of Canada
gbauer@bankofcanada.ca

Antonio Diez de los Rios
Bank of Canada
diez@bankofcanada.ca

August, 2014

A Inter-battery factor analysis

Let $\mathbf{y}_{jt}$ be the vector of $N \times 1$ vector of observable variables for each block (i.e. bond yields in each country) and assume the following joint model for $\mathbf{y}_{jt}$ and $j = 1, \ldots, J$, where J is the number of blocks (i.e. countries):

$$\begin{pmatrix} \mathbf{y}_{1t} \\ \vdots \\ \mathbf{y}_{Jt} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_J \end{pmatrix} + \begin{pmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_J \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} \boldsymbol{\varepsilon}_{1t} \\ \vdots \\ \boldsymbol{\varepsilon}_{Jt} \end{pmatrix} \tag{A.1}$$

$$\begin{pmatrix} \mathbf{f}_t \\ \boldsymbol{\varepsilon}_{1t} \\ \vdots \\ \boldsymbol{\varepsilon}_{Jt} \end{pmatrix} \sim i.i.d.\ N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_{11} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Psi}_{JJ} \end{pmatrix} \right] \tag{A.2}$$

where $\mathbf{f}_t$ is a $K \times 1$ vector of unobserved common factors that are orthogonal to each other, such that $E(\mathbf{f}_t\mathbf{f}_t') = \mathbf{I}_K$ with $\mathbf{I}_K$ a K-dimensional identity matrix; $\mathbf{B}_j$ is a $N \times K$ matrix of constant loadings on the common factors; and $\boldsymbol{\varepsilon}_{jt}$ is a $N \times 1$ vector of idiosyncratic noises with zero, which conditionally orthogonal to $\mathbf{f}_t$ and to any other $\boldsymbol{\varepsilon}_{lt}$ for $l \neq j$. The main difference with respect to a traditional factor model is that we do not assume that the covariance matrix of the idiosyncratic noise for each country ($E(\boldsymbol{\varepsilon}_{jt}\boldsymbol{\varepsilon}_{jt}') = \boldsymbol{\Psi}_{jj}$) is diagonal. Such an assumption implies the potential presence of factors being common to one block only (i.e., a country-specific factor).

Stacking across blocks, let $\mathbf{y}_t = (\mathbf{y}_{1t}', \ldots, \mathbf{y}_{Jt}')'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_J')'$, $\mathbf{B} = (\mathbf{B}_1', \ldots, \mathbf{B}_J')'$, and $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{1t}', \ldots, \boldsymbol{\varepsilon}_{Jt}')'$ so that the model in equations (A.1) and (A.2) can be expressed in compact form as

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t \tag{A.3}$$

$$\begin{pmatrix} \mathbf{f}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} \sim i.i.d.\ N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}\right] \tag{A.4}$$

where $\boldsymbol{\Psi}$ is a $NJ \times NJ$ block diagonal matrix, with relevant block given by $\boldsymbol{\Psi}_{jj}$. These assumptions imply that the distribution of $\mathbf{y}_t$ is multivariate $i.i.d.$ normal with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma} = E(\mathbf{y}_t\mathbf{y}_t') = \mathbf{B}\mathbf{B}' + \boldsymbol{\Psi}$.

Pérignon, Smith and Villa (2007) suggest estimating this common factor model by maximum likelihood. They undertake a numerical maximization of the log-likelihood of a sample of size T of the observable variables:

$$\ln L_T(\boldsymbol{\mu}, \mathbf{B}, \boldsymbol{\Psi}) = \sum_{t=1}^{T} l_t$$

with

$$l_t = -\frac{NJ}{2}\log 2\pi - \frac{1}{2}\ln|\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi}| - \frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu})'(\mathbf{B}\mathbf{B}' + \boldsymbol{\Psi})^{-1}(\mathbf{y}_t - \boldsymbol{\mu})$$

Given the dimension of our problem (four countries and ten maturities per country), such numerical optimization is infeasible. For this reason, we adapt the EM algorithm approach of Sentana (2000), used with conditionally heteroskedastic exact factor models, to our inter-battery factor analysis framework.[1]

The EM algorithm is based on the following identity:

$$\begin{aligned} l(\mathbf{y}_t, \mathbf{f}_t; \boldsymbol{\phi}) &\equiv l(\mathbf{y}_t|\mathbf{f}_t; \boldsymbol{\phi}) + l(\mathbf{f}_t; \boldsymbol{\phi}) \\ l(\mathbf{y}_t, \mathbf{f}_t; \boldsymbol{\phi}) &\equiv l(\mathbf{f}_t|\mathbf{y}_t; \boldsymbol{\phi}) + l(\mathbf{y}_t; \boldsymbol{\phi}) \end{aligned} \tag{A.5}$$

where $l(\mathbf{y}_t, \mathbf{f}_t; \boldsymbol{\phi})$ is the joint log-density function of $\mathbf{y}_t$ and $\mathbf{f}_t$; $l(\mathbf{y}_t|\mathbf{f}_t; \boldsymbol{\phi})$ is the conditional log-density of $\mathbf{y}_t$ given $\mathbf{f}_t$; $l(\mathbf{f}_t; \boldsymbol{\phi})$ the marginal log-density of $\mathbf{f}_t$; $l(\mathbf{f}_t|\mathbf{y}_t; \boldsymbol{\phi})$ the conditional log-density of $\mathbf{f}_t$ given $\mathbf{y}_t$; and $l(\mathbf{y}_t; \boldsymbol{\phi})$ the marginal log-density of $\mathbf{y}_t$, given the parameters vector of parameters of the model, $\boldsymbol{\phi}$. In particular, the EM algorithm exploits the Kullback inequality which states that any increase in $E\left[\sum_t l(\mathbf{y}_t, \mathbf{f}_t|\mathbf{Y}_T; \boldsymbol{\phi})\right]$ must represent an increase in the log-likelihood of the sample $\sum_t l(\mathbf{y}_t|\mathbf{Y}_T; \boldsymbol{\phi})$. The essential steps of this algorithm are the E(stimation)-Step and the M(aximisation)-Step which are carried out at each iteration. So at the n-th iteration we have:

E-Step: Given the current value $\boldsymbol{\phi}^{(n)}$ of the parameter vector and the observed data $\mathbf{Y}_T$, calculate estimates for $\mathbf{f}_t$ as $E(\mathbf{f}_t|\mathbf{Y}_T, \phi^{(n)})$. To this end, notice that the model given by (A.3) and (A.4) has a state space representation where $\mathbf{f}_t$ can be regarded as the state. Under such a representation, (A.3) is the measurement equation given that it describes the relation between the observed variables $\mathbf{y}_t$ and the unobserved factor $\mathbf{f}_t$. The transition equation, on the other hand, has a degenerate nature as it can be represented as $\mathbf{f}_t = \mathbf{0} \cdot \mathbf{f}_{t-1} + \mathbf{f}_t$. We can thus apply the Kalman filter in order to obtain the best (in

[1] Nevertheless, our problem is much more simpler given that we are supposing an $i.i.d.$ sample.

the conditional mean square error sense) estimate of the factor $\mathbf{f}_{t|t} = E(\mathbf{f}_t|\,\mathbf{y}_t)$ and the corresponding mean squared errors $\mathbf{\Omega}_{t|t} = V(\mathbf{f}_t|\,\mathbf{y}_t)$:

$$\mathbf{f}_{t|t} = \mathbf{B}'\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}) \tag{A.6}$$

$$\mathbf{\Omega}_{t|t} = \mathbf{I}_k - \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B} \tag{A.7}$$

Further notice that, given that the transition equation is degenerate, smoothing is unnecesary so that $\mathbf{f}_{t|t} = E(\mathbf{f}_t|\,\mathbf{Y}_T)$, and $\mathbf{\Omega}_{t|t} = V(\mathbf{f}_t|\,\mathbf{Y}_T)$ where $\mathbf{Y}_T = \{\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots\}$.

M-Step: Using the estimated values $\mathbf{f}_{t|t}^{(n)}$ and $\mathbf{\Omega}_{t|t}^{(n)}$, in this step we maximize the expected value of $\sum_t [l(\mathbf{y}_t|\mathbf{f}_t;\boldsymbol{\phi}) + l(\mathbf{f}_t;\boldsymbol{\phi})]$ conditional on $\mathbf{Y}_T$ and the current parameter estimates $\boldsymbol{\phi}^{(n)}$ to determine $\boldsymbol{\phi}^{(n+1)}$. In particular, the objective function at the M-Step of the n-th iteration is:

$$-\frac{TNJ}{2}\log 2\pi - \frac{T}{2}\log|\mathbf{\Psi}| - \frac{1}{2}\sum_{t=1}^{T} tr\left\{\mathbf{\Psi}^{-1}\left[(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_{t|t}^{(n)})(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_{t|t}^{(n)})' + \mathbf{B}\mathbf{\Omega}_{t|t}^{(n)}\mathbf{B}'\right]\right\}$$
$$-\frac{TK}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} tr\left[\mathbf{f}_{t|t}^{(n)}\mathbf{f}_{t|t}^{(n)\prime} + \mathbf{\Omega}_{t|t}^{(n)}\right]$$

and, given our assumptions, this implies that:

$$\widehat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t \tag{A.8}$$

$$\widehat{\mathbf{B}} = \mathbf{B}^{(n+1)} = \left[\sum_{t=1}^{T}\mathbf{f}_{t|t}^{(n)}\mathbf{f}_{t|t}^{(n)\prime} + \mathbf{\Omega}_{t|t}^{(n)}\right]^{-1}\left[\sum_{t=1}^{T}\mathbf{f}_{t|t}^{(n)}(\mathbf{y}_t - \boldsymbol{\mu})'\right]$$

and that $\mathbf{\Psi}_{jj}^{(n+1)}$ can be obtained from the relevant blocks of

$$\frac{1}{T}\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_{t|t}^{(n)})(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_{t|t}^{(n)})' + \mathbf{B}\mathbf{\Omega}_{t|t}^{(n)}\mathbf{B}'\right]$$

Note that, estimates of $\boldsymbol{\mu}$ are independent of the iteration and they coincide with the sample means of $\mathbf{y}_t$. Thus, we can safely apply our inter-battery factor analysis to demeaned data. Furthermore, the expressions obtained in the M-step are very similar to those corresponding to the multivariate regression case and that we would apply to the case in which $\mathbf{f}_t$ were observed. In the M-step, instead, the unobservable factors are replaced by their best (in the conditional mean squared error sense) estimates given the available data.

B Bond Pricing

In this appendix, we show that in a model the state variables, $\mathbf{x}_t$, follow a VAR(1) process:

$$\mathbf{x}_{t+1} = \boldsymbol{\mu} + \mathbf{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}$$

where $\mathbf{v}_t \sim iid\ N(0, \boldsymbol{\Sigma})$, and where the stochastic discount factor (SDF) is given by

$$\xi_{t+1} = \exp(-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1})$$

with a short rate and prices of risk given by the following affine functions:

$$\begin{aligned}\boldsymbol{\lambda}_t &= \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1\mathbf{x}_t \\ r_t &= \delta_0 + \boldsymbol{\delta}_1'\mathbf{x}_t\end{aligned}$$

the log bond prices are affine functions of the state variables

$$p_t^{(n)} = A_n + \mathbf{B}_n'\mathbf{x}_t$$

where A_n and $\mathbf{B}_n$ can be computed recursively:

$$\mathbf{B}_{n+1}' = \mathbf{B}_n'\boldsymbol{\Phi}^Q - \boldsymbol{\delta}_1' \quad \text{(B.1)}$$

$$A_{n+1} = A_n + \mathbf{B}_n'\boldsymbol{\mu}^Q + \frac{1}{2}\mathbf{B}_n'\boldsymbol{\Sigma}\mathbf{B}_n - \delta_0 \quad \text{(B.2)}$$

with $A_1 = -\delta_1$, $\mathbf{B}_1 = -\boldsymbol{\delta}_1$ and $\boldsymbol{\mu}^Q = \boldsymbol{\mu} - \boldsymbol{\lambda}_0$ and $\boldsymbol{\Phi}^Q = \boldsymbol{\Phi} - \boldsymbol{\lambda}_1$ are the matrices governing the dynamic evolution of the state variables under the risk neutral measure.

Note that the affine pricing relationship is trivially satisfied for one-period bonds ($n = 1$) given that

$$p_t^{(1)} = -y_t^{(1)} = -r_t = -\delta_0 - \boldsymbol{\delta}_1'\mathbf{x}_t$$

For $n > 1$, we have that the price at time t of a $n+1$ zero-coupon bond is given by

$$P_t^{(n+1)} = E_t[\xi_{t+1}P_{t+1}^{(n)}]$$

and, thus, we must have

$$\begin{aligned}P_t^{(n+1)} &= E_t\left[\exp\left(-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} + A_n + \mathbf{B}_n'\mathbf{x}_{t+1}\right)\right] \\ &= E_t\left\{\exp\left[-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} + A_n + \mathbf{B}_n'(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1})\right]\right\} \\ &= E_t\left[\exp\left(-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t + A_n + \mathbf{B}_n'\boldsymbol{\mu} + \mathbf{B}_n'\boldsymbol{\Phi}\mathbf{x}_t + (\mathbf{B}_n' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\mathbf{v}_{t+1}\right)\right]\end{aligned}$$

Note, however, that the last term in the previous equation satisfies that

$$\begin{aligned}E_t\left\{\exp\left[(\mathbf{B}_n' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\mathbf{v}_{t+1}\right]\right\} &= \exp\left[\frac{1}{2}(\mathbf{B}_n' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\mathbf{B}_n' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\right] \\ &= \exp\left(\frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \mathbf{B}_n'\boldsymbol{\lambda}_t + \frac{1}{2}\mathbf{B}_n'\boldsymbol{\Sigma}\mathbf{B}_n\right)\end{aligned}$$

Thus we have that

$$\begin{aligned}A_{n+1} + \mathbf{B}_{n+1}'\mathbf{x}_t &= -\delta_0 + A_n + \mathbf{B}_n'\boldsymbol{\mu} - \mathbf{B}_n'\boldsymbol{\lambda}_0 + \frac{1}{2}\mathbf{B}_n'\boldsymbol{\Sigma}\mathbf{B}_n - \boldsymbol{\delta}_1'\mathbf{x}_t + \mathbf{B}_n'\boldsymbol{\Phi}\mathbf{x}_t - \mathbf{B}_n'\boldsymbol{\lambda}_1\mathbf{x}_t \\ &= \left[A_n + \mathbf{B}_n'(\boldsymbol{\mu} - \boldsymbol{\lambda}_0) + \frac{1}{2}\mathbf{B}_n'\boldsymbol{\Sigma}\mathbf{B}_n - \delta_0\right] + \left[\mathbf{B}_n'(\boldsymbol{\Phi} - \boldsymbol{\lambda}_1) - \boldsymbol{\delta}_1'\right]\mathbf{x}_t\end{aligned}$$

And matching coefficients we arrive at the pricing equations above.

C Exchange rates and stochastic discount factors

C.1 Uncovered interest parity under the risk-neutral measure

In this section, we show that the fact that uncovered interest parity must hold under the risk-neutral measure is a direct consequence of the pricing equation of a foreign one-period zero-coupon bond by a domestic investor. In particular, note that the assumption of no-arbitrage implies that the price of a foreign one-period zero-coupon bond, $P_{j,t}^{(1)} = e^{-r_{j,t}}$, must satisfy:

$$E_t\left(\xi_{\$,t+1}e^{\Delta s_{j,t+1}} \times 1/P_{j,t}^{(1)}\right) = E_t\left(\xi_{\$,t+1}e^{\Delta s_{j,t+1}}e^{r_{j,t}}\right) = 1$$

and substituting the law of motion for the rate of depreciation (equation 5 in the main text) and the domestic SDF (equation 6 in the main text) into this last equation

$$\begin{aligned} E_t\left\{\exp\left[-r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} + \mathbf{e}_{F+M+j}'(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}) + r_{j,t}\right]\right\} &= 1 \\ E_t\left\{\exp\left[(r_{j,t} - r_{\$,t}) - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t + \mathbf{e}_{F+M+j}'(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) + (\mathbf{e}_{F+M+j}' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\mathbf{v}_{t+1}\right]\right\} &= 1 \end{aligned}$$

Note, however, that the last term in the previous equation satisfies that

$$\begin{aligned} E_t\left\{\exp\left[(\mathbf{e}_{F+M+j}' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\mathbf{v}_{t+1}\right]\right\} &= \exp\left[\frac{1}{2}(\mathbf{e}_{F+M+j}' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\mathbf{e}_{F+M+j}' - \boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1})\right] \\ &= \exp\left(\frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \mathbf{e}_{F+M+j}'\boldsymbol{\lambda}_t + \frac{1}{2}\mathbf{e}_{F+M+j}'\boldsymbol{\Sigma}\mathbf{e}_{F+M+j}\right) \end{aligned}$$

which implies that

$$(r_{j,t} - r_{\$,t}) + \mathbf{e}_{F+M+j}'(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) - \mathbf{e}_{F+M+j}'\boldsymbol{\lambda}_t + \frac{1}{2}\mathbf{e}_{F+M+j}'\boldsymbol{\Sigma}\mathbf{e}_{F+M+j} = 0.$$

Substituting the expressions for the short-term interest rates and the prices of risk into this equation delivers the following conditions for all $j = 1, ..., J$:

$$\mathbf{e}_j'\boldsymbol{\mu}_3^Q = -\frac{1}{2}\mathbf{e}_j'\boldsymbol{\Sigma}_{33}\mathbf{e}_j + (\delta_\$^{(0)} - \delta_j^{(0)}), \tag{C.1}$$

$$\mathbf{e}_{F+M+j}'\boldsymbol{\Phi}_{3\bullet}^Q = \left[(\boldsymbol{\delta}_\$^{(1)} - \boldsymbol{\delta}_j^{(1)})', \mathbf{0}_M', \mathbf{0}_J'\right]. \tag{C.2}$$

Therefore, we have that uncovered interest parity holds under the risk-neutral measure, Q:

$$E_t^Q\Delta s_{j,t+1} = -\frac{1}{2}Var_t\left(\Delta s_{j,t+1}\right) + (r_{\$,t} - r_{j,t}), \tag{C.3}$$

where $-\frac{1}{2}Var_t\left(\Delta s_{j,t+1}\right)$ is a Jensen's inequality term.

C.2 Affine expected rate of depreciation

We now show that in a model the state variables, $\mathbf{x}_t$, follow a VAR(1) process:

$$\mathbf{x}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}$$

where $\mathbf{v}_t \sim iid\ N(0, \boldsymbol{\Sigma})$, and where the stochastic discount factor (SDF) in the domestic and foreign economy are given by

$$\begin{aligned}
\xi_{\$,t+1} &= \exp(-r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}) \\
\xi_{j,t+1} &= \exp(-r_{j,t} - \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} - \boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1})
\end{aligned}$$

with short rates given by the following affine functions:

$$\begin{aligned}
r_{\$,t} &= \delta_{\$}^{(0)} + \boldsymbol{\delta}'_{\$}\mathbf{x}_t \\
r_{j,t} &= \delta_{j}^{(0)} + \boldsymbol{\delta}'_{j}\mathbf{x}_t
\end{aligned}$$

and prices of risk given by:

$$\begin{aligned}
\boldsymbol{\lambda}_{\$,t} &= \boldsymbol{\lambda}_{\$}^{(0)} + \boldsymbol{\lambda}_{\$}\mathbf{x}_t \\
\boldsymbol{\lambda}_{j,t} &= \boldsymbol{\lambda}_{j}^{(0)} + \boldsymbol{\lambda}_{j}\mathbf{x}_t
\end{aligned}$$

then if the expected rate of depreciation is affine in the set of factors:

$$\Delta s_{j,t} = \gamma_j^{(0)} + \boldsymbol{\gamma}'_j\mathbf{x}_t,$$

the rate of depreciation has to be the difference between the log SDFs in the two countries:

$$\Delta s_{j,t+1} = \log \xi_{j,t+1} - \log \xi_{\$,t+1}. \tag{C.4}$$

We note that when the rate of depreciation is not affine in the factors, an additional assumption of market completeness is needed for equation (C.4) to be a sufficient and necessary condition for exchange rate determination (see Backus, Foresi and Telmer, 2001).

The proof is similar to the one in Anderson, Han and Ramazani (2010) and has three steps. First, since the price of a foreign one-period zero-coupon bond, $P_{j,t}^{(1)} = e^{-r_{j,t}}$, must satisfy:

$$E_t\left(\xi_{\$,t+1}e^{\Delta s_{j,t+1}} \times 1/P_{j,t}^{(1)}\right) = E_t\left(\xi_{\$,t+1}e^{\Delta s_{j,t+1}}e^{r_{j,t}}\right) = 1,$$

we have that the uncovered interest parity hypothesis must hold under the (domestic) risk neutral measure:

$$\gamma_j^{(0)} + \boldsymbol{\gamma}'_j\boldsymbol{\mu}^Q = \frac{1}{2}\boldsymbol{\gamma}'_j\boldsymbol{\Sigma}\boldsymbol{\gamma}_j + \delta_{\$}^{(0)} - \delta_j^{(0)} \tag{C.5}$$

$$\boldsymbol{\gamma}'_j\boldsymbol{\Phi}^Q = \boldsymbol{\delta}'_{\$} - \boldsymbol{\delta}'_j \tag{C.6}$$

where $\boldsymbol{\mu}^Q = \boldsymbol{\mu} - \boldsymbol{\lambda}_{\$}^{(0)}$, and $\boldsymbol{\Phi}^Q = \boldsymbol{\Phi} - \boldsymbol{\lambda}_{j}^{(0)}$.

Second, consider a foreign asset with payoff, $R_{j,t+1}$, at time $t+1$ in units of foreign currency:

$$R_{j,t+1} = \exp\left(r_{j,t} + \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} + \boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}\right).$$

with price given by $E_t\left(\xi_{j,t+1}R_{j,t+1}\right) = E_t\left[\exp(0)\right] = 1$, which implies that $R_{j,t+1}$ is a gross return. Thus, it has to be the case that, when priced by the domestic investor

$$E_t\left(\xi_{\$,t+1}e^{\Delta s_{j,t+1}}R_{j,t+1}\right) = 1.$$

Substituting the law of motion for the rate of depreciation, the domestic SDF, and the expression for the return $R_{j,t+1}$ in the previous expression we get:

$$\begin{aligned} E_t\Big\{\exp\Big[&-r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} + \gamma_j^{(0)} + \boldsymbol{\gamma}'_j(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}) \\ &+r_{j,t} + \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} + \boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}\Big]\Big\} = 1 \\ E_t\Big\{\exp\Big[&(r_{j,t} - r_{\$,t}) - \frac{1}{2}\boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} + \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} + \gamma_j^{(0)} + \boldsymbol{\gamma}'_j(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) \\ &+ \left(\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\gamma}'_j\right)\mathbf{v}_{t+1}\Big]\Big\} = 1 \end{aligned}$$

Note that the last term in the previous equation satisfies that

$$\begin{aligned} &E_t\left\{\exp\left[\left(\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\gamma}'_j\right)\mathbf{v}_{t+1}\right]\right\} = \\ &\exp\left[\frac{1}{2}\left(\boldsymbol{\lambda}_{j,t} - \boldsymbol{\lambda}_{\$,t} + \boldsymbol{\Sigma}\boldsymbol{\gamma}_j\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\lambda}_{j,t} - \boldsymbol{\lambda}_{\$,t} + \boldsymbol{\Sigma}\boldsymbol{\gamma}_j\right)\right] = \\ &\exp\left(\frac{1}{2}\boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} + \frac{1}{2}\boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} + \frac{1}{2}\boldsymbol{\gamma}'_j\boldsymbol{\Sigma}\boldsymbol{\gamma}_j + \boldsymbol{\lambda}'_{j,t}\boldsymbol{\gamma}_j - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\gamma}_j - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t}\right) \end{aligned}$$

Thus we have:

$$\begin{aligned} \exp\Big[&(r_{j,t} - r_{\$,t}) + \boldsymbol{\lambda}'_{j,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} + \gamma_j^{(0)} + \boldsymbol{\gamma}'_j(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) + \frac{1}{2}\boldsymbol{\gamma}'_j\boldsymbol{\Sigma}\boldsymbol{\gamma}_j \\ &+\boldsymbol{\lambda}'_{j,t}\boldsymbol{\gamma}_j - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\gamma}_j - \boldsymbol{\lambda}'_{\$,t}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t}\Big] = 1 \end{aligned}$$

Substituting the expressions for the short rates and the prices of risk and expanding, it can be shown that the previous equation has the following form:

$$\exp\left(A + \mathbf{B}'\mathbf{x}_t + \mathbf{x}'_t\mathbf{C}\mathbf{x}_t\right) = 1$$

which implies that $A = 0$, $\mathbf{B} = \mathbf{0}_{n\times 1}$ and $\mathbf{C} = \mathbf{0}_{n\times n}$ where n is the number of factors in order to $A + \mathbf{B}'\mathbf{x}_t + \mathbf{x}'_t\mathbf{C}\mathbf{x}_t = 0$ for all $\mathbf{x}_t$. In particular, since

$$\mathbf{C} = (\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_\$)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_j,$$

the quadratic term is equal to zero ($\mathbf{C} = \mathbf{0}_{n\times n}$) when

$$\boldsymbol{\lambda}_j = \boldsymbol{\lambda}_\$ = \boldsymbol{\lambda}. \tag{C.7}$$

As for the linear term, we have that

$$\begin{aligned}
\mathbf{B}' &= \boldsymbol{\delta}_j' - \boldsymbol{\delta}_\$' + \boldsymbol{\gamma}_j'\boldsymbol{\Phi} + \left(\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)}\right)' \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} \\
\mathbf{B}' &= \boldsymbol{\gamma}_j'\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^Q\right) - \left(\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)}\right)' \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} \\
\mathbf{B}' &= \boldsymbol{\gamma}_j'\boldsymbol{\lambda} - \left(\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)}\right)' \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} \\
\mathbf{B}' &= \left[\boldsymbol{\gamma}_j' - \left(\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)}\right)' \boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\lambda}
\end{aligned}$$

where the second line comes from the fact that uncovered interest parity is satisfied under Q. Thus the linear term is equal to zero ($\mathbf{B} = \mathbf{0}_{n\times 1}$) when

$$\boldsymbol{\lambda}_j^{(0)} = \boldsymbol{\lambda}_\$^{(0)} - \boldsymbol{\Sigma}\boldsymbol{\gamma}_j. \tag{C.8}$$

Finally, it is possible to show that the constant term:

$$A = (\delta_j^{(0)} - \delta_\$^{(0)}) + \gamma_j^{(0)} + \boldsymbol{\gamma}_j'\boldsymbol{\mu} + (\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_j^{(0)} + (\boldsymbol{\lambda}_j^{(0)} - \boldsymbol{\lambda}_\$^{(0)})'\boldsymbol{\gamma}_j + \frac{1}{2}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}\boldsymbol{\gamma}_j$$

is equal to zero ($A = 0$) under (C.5), (C.6), (C.7) and (C.8).

Third, we show that given this set of restrictions, the rate of depreciation is given by the difference between the log SDFs between the two countries. In particular, we have that

$$\Delta s_{j,t+1} = \gamma_j^{(0)} + \boldsymbol{\gamma}_j'\mathbf{x}_{t+1} = \gamma_j^{(0)} + \boldsymbol{\gamma}_j'(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}).$$

Substituting (C.5) and (C.6) into the previous equation, we arrive at

$$\begin{aligned}
\Delta s_{j,t+1} &= (\delta_\$^{(0)} - \delta_j^{(0)}) + (\boldsymbol{\delta}_\$ - \boldsymbol{\delta}_j)'\mathbf{x}_t + \boldsymbol{\gamma}_j'(\boldsymbol{\mu} - \boldsymbol{\mu}^Q) + \boldsymbol{\gamma}_j'\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^Q\right)\mathbf{x}_t - \frac{1}{2}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}\boldsymbol{\gamma}_j + \boldsymbol{\gamma}_j'\mathbf{v}_{t+1} \\
&= (r_{\$,t} - r_{j,t}) + \boldsymbol{\gamma}_j'(\boldsymbol{\lambda}_\$^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t) - \frac{1}{2}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}\boldsymbol{\gamma}_j + \boldsymbol{\gamma}_j'\mathbf{v}_{t+1}
\end{aligned}$$

Note however that equation (C.8) implies that

$$\begin{gathered}
\boldsymbol{\gamma}_j'(\boldsymbol{\lambda}_\$^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t) - \frac{1}{2}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}\boldsymbol{\gamma}_j = \\
(\boldsymbol{\lambda}_\$^{(0)} - \boldsymbol{\lambda}_j^{(0)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\lambda}_\$^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t) - \frac{1}{2}(\boldsymbol{\lambda}_\$^{(0)} - \boldsymbol{\lambda}_j^{(0)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\lambda}_\$^{(0)} - \boldsymbol{\lambda}_j^{(0)}) = \\
\frac{1}{2}\left(\boldsymbol{\lambda}_\$^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\lambda}_\$^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t\right) - \frac{1}{2}\left(\boldsymbol{\lambda}_j^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\lambda}_j^{(0)} + \boldsymbol{\lambda}\mathbf{x}_t\right) = \\
\frac{1}{2}\boldsymbol{\lambda}_{\$,t}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_{j,t}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t}
\end{gathered}$$

and that

$$\begin{aligned}
\boldsymbol{\gamma}_j'\mathbf{v}_{t+1} &= (\boldsymbol{\lambda}_\$^{(0)} - \boldsymbol{\lambda}_j^{(0)})'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= (\boldsymbol{\lambda}_{\$,t} - \boldsymbol{\lambda}_{j,t})'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}
\end{aligned}$$

Thus we have that:

$$\begin{aligned}
\Delta s_{j,t+1} &= \left(-r_{j,t} - \frac{1}{2}\boldsymbol{\lambda}_{j,t}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{j,t} - \boldsymbol{\lambda}_{j,t}'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}\right) - \left(-r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_{\$,t}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_{\$,t} - \boldsymbol{\lambda}_{\$,t}'\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}\right) \\
&= \log\xi_{j,t+1} - \log\xi_{\$,t+1}
\end{aligned}$$

which is equation (11) in the main text.

C.3 Obtaining the foreign SDF

As noted in the previous section, when the rate of depreciation is affine in the set of pricing factors (which, in our case is trivially satisfied given that $\Delta s_{j,t+1}$ is itself a pricing factor), the law of one price tells us that one of the numeraire SDF, the country j SDF and the rate of depreciation of the currency j is redundant and can be constructed from the other two. In particular, we can solve for the foreign SDF:

$$\log \xi_{j,t+1} = \Delta s_{j,t+1} + \log \xi_{\$,t+1}$$

and substituting the law of motion for the rate of depreciation (equation 5 in the main text) and the domestic SDF (equation 6 in the main text) into this equation and rearranging we obtain:

$$\begin{aligned}
\log \xi_{j,t+1} &= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{v}_{t+1}) - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}'_t\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\lambda}'_t\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}'_t\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t - \left(\boldsymbol{\lambda}'_t - \mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}
\end{aligned}$$

Now define $\boldsymbol{\lambda}_t^{(j)\prime} \equiv \boldsymbol{\lambda}'_t - \mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}$ and use this expression to substitute $\boldsymbol{\lambda}_t$ into the previous equation:

$$\begin{aligned}
\log \xi_{j,t+1} &= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) - r_{\$,t} - \frac{1}{2}\left(\boldsymbol{\lambda}_t^{(j)} + \boldsymbol{\Sigma}\mathbf{e}_{F+M+j}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\lambda}_t^{(j)} + \boldsymbol{\Sigma}\mathbf{e}_{F+M+j}\right) \\
&\quad -\boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_t^{(j)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t^{(j)} - \mathbf{e}'_{F+M+j}\boldsymbol{\lambda}_t^{(j)} \\
&\quad -\frac{1}{2}\mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}\mathbf{e}_{F+M+j} - \boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{x}_t) - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_t^{(j)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t^{(j)} - \mathbf{e}'_{F+M+j}\left(\boldsymbol{\lambda}_t - \boldsymbol{\Sigma}\mathbf{e}_{F+M+j}\right) \\
&\quad -\frac{1}{2}\mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}\mathbf{e}_{F+M+j} - \boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= \mathbf{e}'_{F+M+j}\left[(\boldsymbol{\mu} - \boldsymbol{\lambda}_0) + (\boldsymbol{\Phi} - \boldsymbol{\lambda}_1)\mathbf{x}_t\right] - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_t^{(j)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t^{(j)} \\
&\quad +\frac{1}{2}\mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}\mathbf{e}_{F+M+j} - \boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1} \\
&= \mathbf{e}'_{F+M+j}(\boldsymbol{\mu}^Q + \boldsymbol{\Phi}^Q\mathbf{x}_t) - r_{\$,t} - \frac{1}{2}\boldsymbol{\lambda}_t^{(j)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t^{(j)} + \frac{1}{2}\mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}^{-1}\mathbf{e}_{F+M+j} - \boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}
\end{aligned}$$

Note however that the first term in the previous expression satisfies that:

$$\mathbf{e}'_{F+M+j}(\boldsymbol{\mu}^Q + \boldsymbol{\Phi}^Q\mathbf{x}_t) = E_t^Q\Delta s_{j,t+1} = -\frac{1}{2}\mathbf{e}'_{F+M+j}\boldsymbol{\Sigma}\mathbf{e}_{F+M+j} + (r_{\$,t} - r_{j,t})$$

Thus we have that:

$$\xi_{j,t+1} = \exp\left(-r_{j,t} - \frac{1}{2}\boldsymbol{\lambda}_t^{(j)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_t^{(j)} - \boldsymbol{\lambda}_t^{(j)}\boldsymbol{\Sigma}^{-1}\mathbf{v}_{t+1}\right)$$

D Invariant transformations of multi-country term structure models

Assume the following multi-country term structure model:

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\Delta}_0 + \boldsymbol{\Delta}_1 \mathbf{x}_t \\
\mathbf{x}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi} \mathbf{x}_t + \mathbf{v}_{t+1} \\
\mathbf{x}_{t+1} &= \boldsymbol{\mu}^Q + \boldsymbol{\Phi}^Q \mathbf{x}_t + \mathbf{v}^Q_{t+1}
\end{aligned}$$

where both $\mathbf{v}_t$ and $\mathbf{v}^Q_t$ are *iid* $N(0, \boldsymbol{\Sigma})$, and $\mathbf{x}_t = (\mathbf{x}'_{1,t}, \mathbf{x}'_{2,t})'$ being $\mathbf{x}_{1,t}$ a latent set of factors, and $\mathbf{x}_{2,t}$ observable. As in Dai and Singleton (2000), we interested in applying invariant transformations, $\widehat{\mathbf{x}}_t = \mathbf{c} + \mathbf{D}\mathbf{x}_t$. We then have that the model model above is observationally equivalent to:

$$\begin{aligned}
\mathbf{r}_t &= \widehat{\boldsymbol{\Delta}}_0 + \widehat{\boldsymbol{\Delta}}_1 \mathbf{x}_t \\
\mathbf{x}_{t+1} &= \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\Phi}} \mathbf{x}_t + \widehat{\mathbf{v}}_{t+1} \\
\mathbf{x}_{t+1} &= \widehat{\boldsymbol{\mu}}^Q + \widehat{\boldsymbol{\Phi}}^Q \mathbf{x}_t + \widehat{\mathbf{v}}^Q_{t+1}
\end{aligned}$$

where now both $\widehat{\mathbf{v}}_t$ and $\widehat{\mathbf{v}}^Q_t$ are *iid* $N(0, \widehat{\boldsymbol{\Sigma}})$ and

$$\begin{aligned}
\widehat{\boldsymbol{\Delta}}_0 &= \boldsymbol{\Delta}_0 - \boldsymbol{\Delta}_1 \mathbf{D}^{-1}\mathbf{c} \\
\widehat{\boldsymbol{\Delta}}_1 &= \boldsymbol{\Delta}_1 \mathbf{D}^{-1} \\
\widehat{\boldsymbol{\mu}} &= (\mathbf{I} - \mathbf{D}\boldsymbol{\Phi}\mathbf{D}^{-1})\mathbf{c} + \mathbf{D}\boldsymbol{\mu} \\
\widehat{\boldsymbol{\Phi}} &= \mathbf{D}\boldsymbol{\Phi}\mathbf{D}^{-1} \\
\widehat{\boldsymbol{\mu}}^Q &= (\mathbf{I} - \mathbf{D}\boldsymbol{\Phi}^Q\mathbf{D}^{-1})\mathbf{c} + \mathbf{D}\boldsymbol{\mu}^Q \\
\widehat{\boldsymbol{\Phi}}^Q &= \mathbf{D}\boldsymbol{\Phi}^Q\mathbf{D}^{-1} \\
\widehat{\boldsymbol{\Sigma}} &= \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'
\end{aligned}$$

Of special interest to us are those invariant transformations that leave the set of observable variables, $\mathbf{x}_{2,t}$, unchanged. Such transformations can be expressed the following way:

$$\begin{pmatrix} \widehat{\mathbf{x}}_{1,t} \\ \widehat{\mathbf{x}}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 + \mathbf{D}_1\mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{pmatrix}$$

E Proofs

E.1 Proof of Lemma 1

To proof this lemma, we use the invariant transformations of multi-country term structure models above. In particular, we need to focus on invariant transformations that leave the set of macro variables and exchange rates unchanged:

$$\begin{pmatrix} \widehat{\mathbf{f}}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mathbf{f}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix}.$$

For simplicity, we assume that $\mathbf{\Phi}_{11}^{Q}$ can be diagonalized, that is $\mathbf{\Phi}_{11}^{Q} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}$ where $\mathbf{\Lambda}$ is a diagonal matrix that contains the eigenvalues of $\mathbf{\Phi}_{11}^{Q}$, and $\mathbf{P}$ is a matrix that contains the corresponding eigenvectors. The following two invariant transformations deliver the model in Lemma 1. First, we apply:

$$\begin{pmatrix} \widehat{\mathbf{f}}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} (\mathbf{I}-\mathbf{\Lambda})^{-1}(\mathbf{E}\mathbf{k} - \mathbf{T}^{-1}\boldsymbol{\mu}_1^{Q}) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{T}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mathbf{f}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix}.$$

where

$$\mathbf{E} = \begin{pmatrix} \mathbf{I}_{J+1} \\ \mathbf{0} \end{pmatrix}$$

and $\mathbf{k}$ is a $(J+1)$ dimensional vector.

Second, we exploit that for a given diagonal matrix such as $\mathbf{\Lambda}$, we can pre- and post-multiply it by another diagonal matrix, $\mathbf{B}$, and leave it unchanged it: $\mathbf{\Lambda} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1}$. In particular using

$$\begin{pmatrix} \widetilde{\mathbf{f}}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{L} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{f}}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix}$$

where

$$\mathbf{L} = \begin{pmatrix} \sum_{j=0}^{J} \widehat{\delta}_{1j} & 0 & \dots & 0 \\ 0 & \sum_{j=0}^{J} \widehat{\delta}_{2j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_{j=0}^{J} \widehat{\delta}_{Fj} \end{pmatrix}$$

where $\widehat{\delta}_{ij}$ is the i-th element of vector $\widehat{\boldsymbol{\delta}}_j^{(1)}$, the vector of factor loadings of the short-rate obtained from the first invariant transformation. Under such transformation, the factor loadings for the short-rate will sum up to one, and thus the model can be expressed in the canonical form of Lemma 1 with $\mathbf{\Psi}_{11}^{Q} = \mathbf{\Lambda}$ and $\mathbf{k}_{\infty}^{Q} = \mathbf{k}$.

E.2 Proof of Proposition 2

As noted in the main text, the multi-country term structure model implies yields on domestic and foreign zero coupon bonds that are affine functions of the bond state variables $\mathbf{f}_t$:

$$\mathbf{y}_t = \mathbf{a}_f + \mathbf{b}_f \mathbf{f}_t \tag{E.1}$$

where $\mathbf{y}_t = (y_{\$,t}^{(1)}, \dots, y_{\$,t}^{(N)}, \dots, y_{j,t}^{(n)}, \dots, y_{J,t}^{(N)})'$ and the corresponding elements of $\mathbf{a}_f$ and $\mathbf{b}_f$ are computed using the system of recursive relations in the main text. If we choose to work with "bond" state variables that are linear combinations of the yields themselves, $\mathbf{f}_t = \mathbf{P}'\mathbf{y}_t$, where $\mathbf{P}$ is a full-rank matrix of weights, then, by equation (E.1) we have:

$$\mathbf{f}_t = \mathbf{P}'\mathbf{y}_t = \mathbf{P}'(\mathbf{a}_f + \mathbf{b}_f \mathbf{f}_t),$$

which implies that our model will only be self-consistent when $\mathbf{P}'\mathbf{a}_f = \mathbf{0}_{F\times 1}$ and $\mathbf{P}'\mathbf{b}_f = \mathbf{I}_F$.

On the other hand, the canonical multi-country dynamic term structure model in Lemma 1 implies yields on domestic and foreign zero coupon bonds that are affine in $\mathbf{z}_t$:

$$\mathbf{y}_t = \mathbf{a}_z + \mathbf{b}_z \mathbf{z}_t \tag{E.2}$$

As such, bond state variables are that linear combinations of yields are simply (invariant) affine transformations of the latent factors $\mathbf{z}_t$:

$$\begin{aligned} \mathbf{f}_t &= \mathbf{P}'\mathbf{y}_t \\ &= \mathbf{P}'(\mathbf{a}_z + \mathbf{b}_z \mathbf{z}_t) \\ &= \mathbf{c} + \mathbf{D}\mathbf{z}_t \end{aligned} \tag{E.3}$$

Thus, we can apply the results on invariant transformations above with

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix} = \begin{pmatrix} \mathbf{P}'\mathbf{a}_z \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{P}'\mathbf{b}_z & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mathbf{z}_t \\ \mathbf{m}_t \\ \Delta\mathbf{s}_t \end{pmatrix}$$

to obtain an observationally equivalent model such that:

$$\mathbf{y}_t = \mathbf{a}_z + \mathbf{b}_z \mathbf{z}_t = \mathbf{a}_f + \mathbf{b}_f \mathbf{f}_t$$

Substituting (E.3) into the previous expression:

$$\begin{aligned} \mathbf{a}_z + \mathbf{b}_z \mathbf{z}_t &= \mathbf{a}_f + \mathbf{b}_f \mathbf{f}_t \\ &= \mathbf{a}_f + \mathbf{b}_f \mathbf{P}'\mathbf{y}_t \\ &= \mathbf{a}_f + \mathbf{b}_f (\mathbf{c} + \mathbf{D}\mathbf{z}_t) \\ &= \mathbf{a}_f + \mathbf{b}_f \mathbf{c} + \mathbf{b}_f \mathbf{D}\mathbf{z}_t \end{aligned}$$

Thus, we have that

$$\begin{aligned} \mathbf{b}_z &= \mathbf{b}_f \mathbf{D} \\ \mathbf{a}_z &= \mathbf{a}_f + \mathbf{b}_f \mathbf{c} \end{aligned}$$

Finally, premultiplying by $\mathbf{P}'$, and solving for $\mathbf{P}'\mathbf{b}_f$ in the first equation

$$\mathbf{P}'\mathbf{b}_f = (\mathbf{P}'\mathbf{b}_z)\mathbf{D}^{-1} = \mathbf{I}$$

and solving for $\mathbf{P}'\mathbf{a}_f$ in the second equation

$$\mathbf{P}'\mathbf{a}_f = \mathbf{P}'\mathbf{a}_z - (\mathbf{P}'\mathbf{b}_f)\mathbf{c} = \mathbf{0}$$

we obtain that the corresponding conditions for self-consistency are satisfied.

F Additional details on Step 1: Fitting yields

In this section, we provide additional details on the estimation of the parameters driving the risk-neutral dynamics of the bond factors. In particular, we estimate the parameters under Q directly by minimizing the sum (across maturities, countries, and time) of the squared differences between model predictions and actual yields:

$$\min_{\boldsymbol{\mu}_1^Q, \boldsymbol{\Phi}_{11}^Q, \boldsymbol{\Delta}^{(0)}, \boldsymbol{\Delta}^{(1)}} \sum_{n=1}^{N} \sum_{j=1}^{J+1} \sum_{t=1}^{T} (y_{j,t}^{(n)} - a_j^{(n)} - \mathbf{b}_j^{(n)\prime} \mathbf{f}_t)^2. \tag{F.1}$$

subject to the self-consistency restrictions in Proposition 2 in the main text. As JPS, we focus on the case where the eigenvalues in $\boldsymbol{\Psi}_{11}^Q$ are real and distinct. In order to satisfy Hamilton and Wu's (2012) identification restriction (see Remark 1 of Lemma 1), we assume that the eigenvalues of $\boldsymbol{\Psi}_{11}^Q$ are distributed according to a power law relation:

$$\psi_{11,j}^Q = \overline{\psi}_{11}^Q \varphi^{j-1}, \qquad j = 1, \ldots, F \tag{F.2}$$

where $\overline{\psi}_{11}^Q$ is the largest eigenvalue of $\boldsymbol{\Psi}_{11}^Q$ and the power scaling coefficient $0 < \varphi < 1$ controls the spacing between different eigenvalues. We refer the reader to Calvet, Fisher and Wu (2013) for an application of power law structures to term structure modeling. Moreover, as in Christensen, Diebold and Rudebusch (2010), we set $\overline{\psi}_{11}^Q = 1.00$ in order to replicate the level factor that characterizes the international cross-section of interest rates. In this way, we reduce the number of free parameters to 29 and we estimate them by minimizing (F.1) directly.

We estimate the power scaling coefficient, φ, sequentially through concentration. That is, for a given value of φ, we numerically minimize the sum of the squared differences between model predictions and actual yields as a function of the rest of parameters driving the risk-neutral measure. We then search over the possible values of φ for the one that minimizes the sum of squared differences to get our estimate of this parameter.

Below, we show that it is possible to concentrate out $\mathbf{k}_\infty^Q$ which further reduces the number of free parameters to be estimated directly in the minimization of (F.1).

We also employ a score algorithm to minimize the sum of squared differences between actual and model-implied yields, with analytical expressions for the gradient and the expected value of the Hessian of the criterion function. By providing both the gradient and an estimate of the Hessian of the criterion function, we obtain a very fast convergence of our optimization algorithm (e.g., around one minute for an eight factor and four country model).

Finally, we note that we could have chosen to work with linear combinations of yields that resemble empirical measures of level, slope and curvature. For example, we could define the level as the 10-year yield, the slope as the difference between the 10- and 1-year yields, and the curvature as twice the 5-year yield minus the sum of the 1- and 10-year yields. However, in our estimation below we assume that the bond state factors are priced perfectly by our model. Thus, by using principal components, we account for as much of the variability in the international cross-section of yields as possible, which, in turns, greatly minimizes the pricing errors of the model with respect to any other linear combination of yields that could potentially be used.

F.1 Concentrating $\mathbf{k}_\infty^Q$ out

In order to concentrate $\mathbf{k}_\infty^Q$ out of the sum of squared residuals, first note that the canonical specification of our multi-country term structure model implies:

$$y_{j,t}^{(n)} = a_{j,z}^{(n)} + \mathbf{b}_{j,z}^{(n)}\mathbf{z}_t \tag{F.3}$$

where $b_{j,z}^{(n)}$ does not depend on $\mathbf{k}_\infty^Q$ and

$$\begin{aligned} a_{j,z}^{(n)} &= \boldsymbol{\alpha}_{j,z}\mathbf{k}_\infty^Q + \boldsymbol{\beta}_{j,z}\boldsymbol{\Omega}_{13}\mathbf{e}_j + \boldsymbol{\gamma}_{j,z}vec(\boldsymbol{\Omega}_{11}) \qquad \text{for } j = 1, ..., J \\ a_{j,z}^{(n)} &= \boldsymbol{\alpha}_{j,z}\mathbf{k}_\infty^Q + \boldsymbol{\gamma}_{j,z}vec(\boldsymbol{\Omega}_{11}) \qquad \text{for } j = \$ \end{aligned}$$

where

$$\boldsymbol{\beta}_{j,z} = \begin{pmatrix} 0 \\ \frac{1}{2}\mathbf{b}_{j,z}^{(1)} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n-1} i\mathbf{b}_{j,z}^{(i)} \end{pmatrix} \qquad \boldsymbol{\gamma}_{j,z} = \begin{bmatrix} 0 \\ -\frac{1}{2} \times \frac{1}{2}\left(\mathbf{b}_{j,z}^{(1)\prime} \otimes \mathbf{b}_{j,z}^{(1)\prime}\right) \\ \vdots \\ -\frac{1}{n} \times \frac{1}{2}\sum_{i=1}^{n-1} i^2 \left(\mathbf{b}_{j,z}^{(i)\prime} \otimes \mathbf{b}_{j,z}^{(i)\prime}\right) \end{bmatrix}$$

and $\boldsymbol{\alpha}_{j,z} = \boldsymbol{\beta}_{j,z}\mathbf{E}$ with

$$\mathbf{E} = \begin{pmatrix} \mathbf{I}_{J+1} \\ \mathbf{0} \end{pmatrix}$$

Second, stacking the pricing equations across countries and maturities, we obtain:

$$\mathbf{y}_t = \mathbf{a}_z + \mathbf{b}_z\mathbf{z}_t \tag{F.4}$$

where, again, $\mathbf{b}_z$ does not depend on $\mathbf{k}_\infty^Q$ and

$$\mathbf{a}_z = \boldsymbol{\alpha}_z\mathbf{k}_\infty^Q + \boldsymbol{\beta}_z\widetilde{\boldsymbol{\theta}}^Q + \boldsymbol{\gamma}_z vec(\boldsymbol{\Omega}_{11}) \tag{F.5}$$

with

$$\boldsymbol{\alpha}_z = \begin{pmatrix} \boldsymbol{\alpha}_{\$,z} \\ \boldsymbol{\alpha}_{1,z} \\ \vdots \\ \boldsymbol{\alpha}_{J,z} \end{pmatrix} \qquad \boldsymbol{\beta}_z = \begin{pmatrix} \boldsymbol{\beta}_{\$,z} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\beta}_{1,z} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\beta}_{J,z} \end{pmatrix} \qquad \boldsymbol{\gamma}_z = \begin{pmatrix} \boldsymbol{\gamma}_{\$,z} \\ \boldsymbol{\gamma}_{1,z} \\ \vdots \\ \boldsymbol{\gamma}_{J,z} \end{pmatrix}$$

and

$$\widetilde{\boldsymbol{\theta}}^Q = \begin{bmatrix} \mathbf{0} \\ vec(\boldsymbol{\Omega}_{13}) \end{bmatrix}$$

Third, rotating the factors to $\mathbf{f}_t = \mathbf{P}'\mathbf{y}_t$, we have that

$$\begin{gathered} \mathbf{f}_t = \mathbf{P}'\mathbf{y}_t = \mathbf{P}'(\mathbf{a}_z + \mathbf{b}_z\mathbf{z}_t) = \mathbf{c} + \mathbf{D}\mathbf{z}_t \\ \boldsymbol{\Sigma}_{13} = \mathbf{D}\boldsymbol{\Omega}_{13} \longrightarrow vec(\boldsymbol{\Omega}_{13}) = (\mathbf{I}_J \otimes \mathbf{D}^{-1})vec(\boldsymbol{\Sigma}_{13}) \\ \boldsymbol{\Sigma}_{11} = \mathbf{D}\boldsymbol{\Omega}_{13}\mathbf{D}' \longrightarrow vec(\boldsymbol{\Omega}_{11}) = (\mathbf{D}^{-1} \otimes \mathbf{D}^{-1})vec(\boldsymbol{\Sigma}_{11}) \\ \widetilde{\boldsymbol{\theta}}^Q = (\mathbf{I}_{J+1} \otimes \mathbf{D}^{-1})\widetilde{\boldsymbol{\mu}}^Q \end{gathered}$$

where

$$\widetilde{\boldsymbol{\mu}}^Q = \begin{bmatrix} \mathbf{0} \\ vec(\boldsymbol{\Sigma}_{13}) \end{bmatrix}$$

Therefore, we can write the following model the bond yields as a function of the new factors $\mathbf{f}_t$:

$$\mathbf{y}_t = \mathbf{a}_f + \mathbf{b}_f \mathbf{f}_t$$

where, $\mathbf{b}_f$ does not depend on $\mathbf{k}_\infty^Q$ and

$$\mathbf{a}_f = \boldsymbol{\alpha}_f \mathbf{k}_\infty^Q + \boldsymbol{\beta}_f \widetilde{\boldsymbol{\mu}}^Q + \boldsymbol{\gamma}_f vec(\boldsymbol{\Sigma}_{11}) \tag{F.6}$$

with

$$\begin{aligned} \mathbf{F} &= \mathbf{I} - (\mathbf{b}_z \mathbf{D}^{-1})\mathbf{P}' \\ \boldsymbol{\alpha}_f &= \mathbf{F}\boldsymbol{\alpha}_z \\ \boldsymbol{\beta}_f &= \mathbf{F}\boldsymbol{\beta}_z(\mathbf{I}_{J+1} \otimes \mathbf{D}^{-1}) \\ \boldsymbol{\gamma}_f &= \mathbf{F}\boldsymbol{\gamma}_z(\mathbf{D}^{-1} \otimes \mathbf{D}^{-1}) \end{aligned}$$

Finally, note that equation (F.6) is linear in $\mathbf{k}_\infty^Q$. Therefore, when solving the first order condition of the optimization problem with respect to $\mathbf{k}_\infty^Q$ we have that

$$\begin{aligned} SSR &= \sum_{t=1}^{T} \mathbf{u}_t' \mathbf{u}_t \\ \frac{\partial SSR}{\partial \mathbf{k}_\infty^Q} &= \sum_{t=1}^{T} \frac{\partial \mathbf{u}_t'}{\partial \mathbf{k}_\infty^Q} \mathbf{u}_t = 0 \end{aligned}$$

where

$$\begin{aligned} \mathbf{u}_t = \mathbf{y}_t - \boldsymbol{\alpha}_f \mathbf{k}_\infty^Q - \boldsymbol{\beta}_f \widetilde{\boldsymbol{\mu}}^Q - \boldsymbol{\gamma}_f vec(\boldsymbol{\Sigma}_{11}) - \mathbf{b}_f \mathbf{f}_t \\ \frac{\partial \mathbf{u}_t'}{\partial \mathbf{k}_\infty^Q} = -\boldsymbol{\alpha}_f' \end{aligned}$$

which implies that $\widehat{\mathbf{k}}_\infty^Q$ must satisfy:

$$\widehat{\mathbf{k}}_\infty^Q = (\boldsymbol{\alpha}_f' \boldsymbol{\alpha}_f)^{-1} \left\{ \boldsymbol{\alpha}_f' \left[\overline{\mathbf{y}} - \boldsymbol{\beta}_f \widetilde{\boldsymbol{\mu}}^Q - \boldsymbol{\gamma}_f vec(\boldsymbol{\Sigma}_{11}) - \mathbf{b}_f \overline{\mathbf{f}} \right] \right\}$$

where $\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t$ and $\overline{\mathbf{f}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}_t$.

F.2 Details on the optimization algorithm

To speed up the minimization of the sum (across maturities, countries, and time) of squared differences between actual and model-implied yields, we use a scoring algorithm, which is a Newton-Raphson optimization algorithm where one approxi mates the Hessian of the function to be minimized by its expectation.

In particular, let $f(\boldsymbol{\psi})$ be the function to be minimized

$$f(\boldsymbol{\psi}) = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{j=1}^{J+1}(y_{j,t}^{(n)} - a_j^{(n)} - \mathbf{b}_j^{(n)\prime}\mathbf{f}_t)^2 = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{j=1}^{J+1}\left(u_{j,t}^{(n)}\right)^2$$

where $\boldsymbol{\psi} = \left[\mathbf{k}_\infty^{Q\prime}, vec(\boldsymbol{\Gamma})'\right]'$ is the vector of structural parameters. Then, the relevant elements of the gradient vector and hessian matrix of $f(\boldsymbol{\psi})$ can be obtained from:

$$g_k = \frac{\partial f(\boldsymbol{\psi})}{\partial \psi_k} = 2\sum_{n=1}^{N}\sum_{j=1}^{J+1}\sum_{t=1}^{T}\frac{\partial u_{j,t}^{(n)}}{\partial \psi_k}u_{j,t}^{(n)}$$

$$h_{kl} = \frac{\partial^2 f(\boldsymbol{\psi})}{\partial \psi_k \partial \psi_l} = 2\sum_{n=1}^{N}\sum_{j=1}^{J+1}\sum_{t=1}^{T}\left[\frac{\partial^2 u_{j,t}^{(n)}}{\partial \psi_k \partial \psi_l}u_{j,t}^{(n)} + \frac{\partial u_{j,t}^{(n)}}{\partial \psi_k}\frac{\partial u_{j,t}^{(n)}}{\partial \psi_l}\right]$$

The idea of the scoring algorithm is to replace the true h_{kl} by the approximate hessian that does not depend on second derivatives:

$$\widetilde{h}_{kl} = \frac{\partial^2 f(\boldsymbol{\psi})}{\partial \psi_k \partial \psi_l} = 2\sum_{n=1}^{N}\sum_{j=1}^{J+1}\sum_{t=1}^{T}\left[\frac{\partial u_{j,t}^{(n)}}{\partial \psi_k}\frac{\partial u_{j,t}^{(n)}}{\partial \psi_l}\right]$$

This choice can be justified under the assumption that pricing errors $u_{j,t}^{(n)}$ are orthogonal to the pricing factors, $\mathbf{f}_t$. In such a case, when T is sufficiently large, the first term of h_{kl} vanishes:

$$\sum_{n=1}^{N}\sum_{j=1}^{J+1}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 u_{j,t}^{(n)}}{\partial \psi_k \partial \psi_l}u_{j,t}^{(n)}\right] \to \sum_{n=1}^{N}\sum_{j=1}^{J+1}E\left[\frac{\partial^2 u_{j,t}^{(n)}}{\partial \psi_k \partial \psi_l}u_{j,t}^{(n)}\right] = 0$$

given that $\partial^2 u_{j,t}^{(n)}/\partial \psi_k \partial \psi_l$ is a function of $\mathbf{f}_t$.

In turn, both the gradient and (approximate) hessian of the sum of squared residuals require the analytical derivatives of the pricing errors with respect to the structural parameters:

$$\frac{\partial u_{j,t}^{(n)}}{\partial \psi_k} = -\frac{\partial a_j^{(n)}}{\partial \psi_k} - \frac{\partial \mathbf{b}_j^{(n)\prime}}{\partial \psi_k}\mathbf{f}_t$$

and, thus, the derivatives of the bond price coefficients $a_j^{(n)} = -A_j^{(n)}/n$ and $\mathbf{b}_j^{(n)} = -\mathbf{B}_j^{(n)}/n$ which we can evaluate analytically by an extra set of recursions that run in parallel with the pricing equations. As shown in Diez de los Rios (2010), these extra recursions are obtained by differentiating the pricing equations in (B.1) and (B.2). For example for the case of the numeraire currency:

$$\frac{\partial A_\$^{(n)}}{\partial \psi_i} = \frac{\partial A_\$^{(n-1)}}{\partial \psi_i} + \frac{\partial \mathbf{B}_\$^{(n-1)\prime}}{\partial \psi_i}\boldsymbol{\mu}_1^Q + \mathbf{B}_\$^{(n-1)\prime}\frac{\partial \boldsymbol{\mu}_1^Q}{\partial \psi_i} + \frac{\partial \mathbf{B}_\$^{(n-1)\prime}}{\partial \psi_i}\boldsymbol{\Sigma}_{11}\mathbf{B}_\$^{(n-1)}$$
$$+\frac{1}{2}\mathbf{B}_\$^{(n-1)\prime}\frac{\partial \boldsymbol{\Sigma}_{11}}{\partial \psi_i}\mathbf{B}_\$^{(n-1)} - \frac{\partial \delta_\$^{(0)}}{\partial \psi_i},$$
$$\frac{\partial \mathbf{B}_\$^{(n)\prime}}{\partial \psi_i} = \frac{\partial \mathbf{B}_\$^{(n-1)\prime}}{\partial \psi_i}\boldsymbol{\Phi}_{11}^Q + \mathbf{B}_\$^{(n-1)\prime}\frac{\partial \boldsymbol{\Phi}_{11}^Q}{\partial \psi_i} - \frac{\partial \boldsymbol{\delta}_\$^{(1)\prime}}{\partial \psi_i}.$$

with $\partial A_\$^{(1)}/\partial \psi_i = -\partial \delta_\$^{(0)}/\partial \psi_i$ and $\partial \mathbf{B}_\$^{(1)}/\partial \psi_i = -\partial \boldsymbol{\delta}_\$^{(1)}/\partial \psi_i$.

G Standard Errors

In this appendix, we provide standard errors for the estimation of the parameters driving the dynamics of the pricing factors under the risk neutral measure, and those driving the price of bond and foreign exchange risk.

Stage 1. Remember that we estimate the parameters governing the risk-neutral distribution Q, $\boldsymbol{\psi} = \left[\mathbf{k}_{\infty}^{Q\prime}, vec(\boldsymbol{\Gamma})'\right]'$, by minimizing the sum (across maturities, countries and time) of the squared differences between model predictions and actual yields:

$$\widehat{\boldsymbol{\psi}} = \arg\min f(\boldsymbol{\psi}) = \arg\min \sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{j=1}^{J+1} \left(u_{j,t}^{(n)}\right)^2$$

where $u_{j,t}^{(n)} = y_{j,t}^{(n)} - a_j^{(n)} - \mathbf{b}_j^{(n)\prime}\mathbf{f}_t$. Since dividing the criterion function by T does not change the solution to our minimization problem, we can think of $\widehat{\boldsymbol{\psi}}$, as the solution to the sample analog of the following set of moment conditions:

$$E\left[2\sum_{n=1}^{N}\sum_{j=1}^{J+1} \frac{\partial u_{j,t}^{(n)}}{\partial \psi_i} u_{j,t}^{(n)}\right] = E\left[s_{it}^{(1)}\right] = 0 \qquad \forall i$$

Then, we can use standard GMM asymptotic theory to obtain standard errors for $\widehat{\boldsymbol{\psi}}$:

$$\sqrt{T}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \rightarrow N\left[\mathbf{0}, \left(\mathbf{D}_{11}'\mathbf{S}_{11}^{-1}\mathbf{D}_{11}\right)^{-1}\right]$$

where $\mathbf{D}_{11} = E\left[\partial \mathbf{s}_t^{(1)}/\partial \boldsymbol{\psi}'\right]$ and $\mathbf{S}_{11} = \sum_{j=-\infty}^{\infty} E\left[\mathbf{s}_t^{(1)}\mathbf{s}_{t-j}^{(1)\prime}\right]$. Moreover, we can use the results in the previous appendix to show that, under the assumption that $u_{j,t}^{(n)}$ is orthogonal to the bond pricing factors, $\mathbf{D}_{11}$ only depends on the first derivatives of the bond price coefficients $a_j^{(n)} = -A_j^{(n)}/n$ and $\mathbf{b}_j^{(n)} = -\mathbf{B}_j^{(n)}/n$, which greatly simplifies obtaining an algebraic expression for the estimate of $\mathbf{D}_{11}$.

Stage 2. The parameters driving the price of bond and foreign exchange risks are, on the other hand, obtained from OLS regressions on the bond pricing factors:

$$\begin{aligned}
\mathbf{f}_{t+1} - \left(\widehat{\boldsymbol{\theta}}_1^Q + \widehat{\boldsymbol{\Phi}}_{11}^Q\mathbf{f}_t\right) &= \boldsymbol{\lambda}_{10} + \boldsymbol{\lambda}_{11}\mathbf{f}_t + \boldsymbol{\lambda}_{12}\mathbf{m}_t + \boldsymbol{\lambda}_{13}\Delta\mathbf{s}_t + \mathbf{v}_{1,t+1} \\
\Delta\mathbf{s}_{t+1} - \left(\widehat{\boldsymbol{\theta}}_3^Q + \widehat{\boldsymbol{\Phi}}_{31}^Q\mathbf{f}_t\right) &= \boldsymbol{\lambda}_{30} + \boldsymbol{\lambda}_{31}\mathbf{f}_t + \boldsymbol{\lambda}_{32}\mathbf{m}_t + \boldsymbol{\lambda}_{33}\Delta\mathbf{s}_t + \mathbf{v}_{3,t+1}
\end{aligned}$$

where $\widehat{\boldsymbol{\theta}}_1^Q$, $\widehat{\boldsymbol{\theta}}_3^Q$, $\widehat{\boldsymbol{\Phi}}_{11}^Q$ and $\widehat{\boldsymbol{\Phi}}_{31}^Q$ are estimates of the parameters under the risk-neutral measure obtained in the first stage.[2]

Thus, we could potentially obtain standard errors once we recast our estimation within the GMM framework using the moment conditions that are implicit in the OLS estimation:

$$E\begin{pmatrix} \mathbf{v}_{1,t+1} \otimes \mathbf{x}_t \\ \mathbf{v}_{3,t+1} \otimes \mathbf{x}_t \end{pmatrix} = E\left[\mathbf{s}_t^{(2)}\right] = 0$$

[2]To keep notation simple, we focus here on the case of unrestricted prices of risk.

where $\mathbf{x}_t = (\mathbf{f}_t', \mathbf{m}_t', \Delta\mathbf{s}_t')'$. However, inference will not be valid in this context because it doesn't take into account that $\mathbf{s}_t^{(2)}$ depends not only on the coefficients driving the price of risk, $\boldsymbol{\lambda}$, but also on the parameters governing the risk-neutral distribution, $\boldsymbol{\psi}$. In order to correct for this "generated regressors problem," we simply stack the moment conditions corresponding to both stages:

$$E\begin{bmatrix} \mathbf{s}_t^{(1)}(\boldsymbol{\psi}) \\ \mathbf{s}_t^{(2)}(\boldsymbol{\psi}, \boldsymbol{\lambda}) \end{bmatrix} = E\left[\mathbf{s}_t(\boldsymbol{\psi}, \boldsymbol{\lambda})\right] = 0$$

and then use standard asymptotic theory to obtain

$$\sqrt{T}(\left[\begin{pmatrix} \widehat{\boldsymbol{\psi}} \\ \widehat{\boldsymbol{\lambda}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\lambda} \end{pmatrix}\right] \rightarrow N\left[\mathbf{0}, \left(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D}\right)^{-1}\right]$$

where

$$\mathbf{D} = E\left[\frac{\partial \mathbf{s}_t}{\partial(\boldsymbol{\psi}', \boldsymbol{\lambda}')'}\right] = E\begin{bmatrix} \partial\mathbf{s}_t^{(1)}/\partial\boldsymbol{\psi}' & \mathbf{0} \\ \partial\mathbf{s}_t^{(2)}/\partial\boldsymbol{\psi}' & \partial\mathbf{s}_t^{(2)}/\partial\boldsymbol{\lambda}' \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix}$$

and $\mathbf{S} = \sum_{j=-\infty}^{\infty} E\left[\mathbf{s}_t \mathbf{s}_{t-j}{}'\right]$

H Domestic Asset Pricing

In this appendix, we focus on the set of individual models for each country where (i) we allow for non-zero prices of risk for each country's two domestic principal components, and (ii) time-variation in the prices of risk is only driven by domestic factors. In particular, we show how to cast this collection of domestic princing models into our multi-country framework in order to compare the implied Sharpe ratios of domestic and international asset pricing models.

First note that we can represent the model for each country in terms of our canonical representation in Lemma 1 under appropriate zero restrictions on $\boldsymbol{\Gamma}^{(1)}$. In particular, we have the matrix of short-rate factor loadings:

$$\begin{pmatrix} r_{\$,t} \\ r_{1,t} \\ \vdots \\ r_{J,t} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_L' & 0 & \cdots & 0 \\ 0 & \mathbf{1}_L' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_L' \end{pmatrix} \begin{pmatrix} \widetilde{\mathbf{z}}_{\$,t} \\ \widetilde{\mathbf{z}}_{1,t} \\ \vdots \\ \widetilde{\mathbf{z}}_{J,t} \end{pmatrix} \tag{H.1}$$

$$\mathbf{r}_t = \boldsymbol{\Gamma}^{(1)}\widetilde{\mathbf{z}}_t$$

where L is the number of domestic bond factors per country, $\widetilde{\mathbf{z}}_{j,t}$ is a vector that collects the set of domestic bond factors for country j, and $\mathbf{1}_L$ is a L-dimensional vector of ones.

On the other hand, the dynamics of the domestic latent factors are given by

$$\begin{pmatrix} \widetilde{z}_{j1,t} \\ \widetilde{z}_{j2,t} \\ \vdots \\ \widetilde{z}_{jL,t} \end{pmatrix} = \begin{pmatrix} k_{j,\infty}^Q \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \psi_{11,j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{11,j}^{L-1} \end{pmatrix} \begin{pmatrix} \widetilde{z}_{j1,t-1} \\ \widetilde{z}_{j2,t-1} \\ \vdots \\ \widetilde{z}_{jL,t-1} \end{pmatrix} + \begin{pmatrix} u_{j1,t}^Q \\ u_{j2,t}^Q \\ \vdots \\ u_{jL,t}^Q \end{pmatrix} \tag{H.2}$$

$$\widetilde{\mathbf{z}}_{j,t} = \mathbf{e}_1 k_{j,\infty}^Q + \boldsymbol{\Psi}_{11,j}^Q \mathbf{z}_{j,t-1} + \mathbf{u}_{j1,t}^Q$$

In particular, we have assumed that $\widetilde{\mathbf{z}}_{j,t}$ follows an autonomous VAR(1) process under the risk-neutral measure and that the eigenvalues of $\mathbf{\Psi}^{Q}_{11,j}$ are distributed according to a power law relation. Again, in order to replicate a domestic level factor for each one of the countries, we set the largest eigenvalue of $\mathbf{\Psi}^{Q}_{11,j}$ to 1.00.

The joint dynamics of $\widetilde{\mathbf{z}}_t = (\widetilde{\mathbf{z}}_{1,t}, \widetilde{\mathbf{z}}_{2,t}, ..., \widetilde{\mathbf{z}}_{J,t})'$ can thus be cast in terms of the canonical representation in Lemma 1 in the main text with appropriate restrictions.

Then, we choose domestic state variables that are linear combinations of the domestic yields only:

$$\widetilde{\mathbf{f}}_{j,t} = \widetilde{\mathbf{P}}_j \mathbf{y}_{j,t} \tag{H.3}$$

where $\mathbf{y}_{j,t} = (y^{(1)}_{j,t}, \ldots, y^{(N)}_{j,t})'$ is a vector that collects all the yields for a given country and $\widetilde{\mathbf{P}}_j$ is a full-rank matrix of weights. We choose the first $L = 2$ principal components cross-section of yields for a given country. Stacking the domestic factors for each country, we find that, under domestic asset pricing, we also have bond factors that are linear combinations of the yields themselves.

$$\begin{pmatrix} \widetilde{\mathbf{f}}_{\$,t} \\ \widetilde{\mathbf{f}}_{1,t} \\ \vdots \\ \widetilde{\mathbf{f}}_{J,t} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{P}}_{\$} & 0 & \cdots & 0 \\ 0 & \widetilde{\mathbf{P}}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{\mathbf{P}}_J \end{pmatrix} \begin{pmatrix} \mathbf{y}_{1,t} \\ \mathbf{y}_{2,t} \\ \vdots \\ \mathbf{y}_{J,t} \end{pmatrix}$$

$$\widetilde{\mathbf{f}}_t = \widetilde{\mathbf{P}} \mathbf{y}_t$$

Therefore, we can use the results in Proposition 2 in the main text to obtain the self-consistency restrictions implied by the domestic pricing.

Finally, we assume the prices of risk are affine and that time-variation in the prices of risk is only driven by domestic factors, that is, domestic bond factors and domestic macroeconomic variables:

$$\boldsymbol{\lambda}_{j,t} = \boldsymbol{\lambda}_{j0} + \boldsymbol{\lambda}_{j1} \widetilde{\mathbf{f}}_{j,t} + \boldsymbol{\lambda}_{j2} \mathbf{m}_{j,t} \tag{H.4}$$

where $\mathbf{m}_{j,t} = (g_{j,t}, \pi_{j,t})'$ is a vector that collects country j's growth and inflation rates. We can stack (H.4) for each country to obtain that

$$\begin{pmatrix} \boldsymbol{\lambda}_{\$,t} \\ \boldsymbol{\lambda}_{1,t} \\ \vdots \\ \boldsymbol{\lambda}_{J,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}_{\$0} \\ \boldsymbol{\lambda}_{10} \\ \vdots \\ \boldsymbol{\lambda}_{J0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}_{\$1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda}_{11} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda}_{J1} \end{pmatrix} \begin{pmatrix} \widetilde{\mathbf{f}}_{\$,t} \\ \widetilde{\mathbf{f}}_{1,t} \\ \vdots \\ \widetilde{\mathbf{f}}_{J,t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}_{\$2} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda}_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda}_{J2} \end{pmatrix} \begin{pmatrix} \mathbf{m}_{\$,t} \\ \mathbf{m}_{1,t} \\ \vdots \\ \mathbf{m}_{J,t} \end{pmatrix}$$

One can thus understand the domestic asset pricing model as a multi-country model where zero restrictions are imposed on the general characterization of the prices of risk $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda} \mathbf{x}_t$.

References

[1] Anderson, B., P.J. Hammond and C.A. Ramezani (2010): "Affine Models of the Joint Dynamics of Exchange Rates and Interest Rates," *Journal of Financial and Quantitative Analysis* 45, 1341-1365.

[2] Backus, D.K., S. Foresi and C.I. Telmer (2001): "Affine Term Structure Models and the Forward Premium Anomaly," *Journal of Finance*, 51, 279-304.

[3] Calvet L.E., A. Fisher and L. Wu (2013): "Staying on Top of the Curve: A Cascade Model of Term Structure Dynamics", mimeo University of British Columbia.

[4] Christensen, J., F.X. Diebold and G. Rudebusch (2011): "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models, " *Journal of Econometrics* 164, 4-20.

[5] Dai, Q. and K.J. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943-1978.

[6] Diez de los Rios, A. (2010): "McCallum Rules, Exchange Rates, and the Term Structure of Interest Rates," mimeo, Bank of Canada.

[7] Hamilton, J.D. and J.C. Wu (2012): "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168, 315-331.

[8] Perignon C., D.R. Smith and C. Villa (2007): "Why common factors in international bond returns are not so common," *Journal of International Money and Finance*, 26, 284-304.

[9] Sentana E. (2000): "The Likelihood Function of Conditionally Heteroskedastic Factor Models," *Annales d'Economie et de Statistique* ,58, 1-19.