

Online Appendices

A Bond Prices Under Risk-Neutral Dynamics

Substituting the exponential-affine solution for bond prices into the no-arbitrage condition yields

$$\begin{aligned}\exp(-a_{s_t}(\tau) - b(\tau)'f_t) &= P_t(\tau) = \mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t)P_{t+1}(\tau-1)] \\ &= \mathbb{E}_t^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t)P_{t+1}(\tau-1) \mid s_{t+1}]],\end{aligned}$$

where the second equality is justified by the law of iterated expectations. Note that

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t)P_{t+1}(\tau-1) \mid s_{t+1}] &= \mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t - a_{s_{t+1}}(\tau-1) - b(\tau-1)'f_{t+1}) \mid s_{t+1}] \\ &= \exp(-r_t - a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} - b(\tau-1)'G^{\mathbb{Q}}f_t) \\ &\quad \times \mathbb{E}_t^{\mathbb{Q}}[\exp(-b(\tau-1)'\Omega_{s_{t+1}}^{\frac{1}{2}}v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}].\end{aligned}$$

By the formula for the moment-generating function of multivariate standard normally distributed random vectors and the fact that $v_{t+1}^{\mathbb{Q}} \mid s_{t+1} \sim \mathcal{N}(O_{3 \times 1}, I_3)$ under $\mathbb{Q}$, we have

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}[\exp(-b(\tau-1)'\Omega_{s_{t+1}}^{\frac{1}{2}}v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}] &= \exp\left(\frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}^{\frac{1}{2}}\Omega_{s_{t+1}}^{\frac{1}{2}'}b(\tau-1)\right) \\ &= \exp\left(\frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right).\end{aligned}$$

As such,

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}[\exp(-r_t)P_{t+1}(\tau-1) \mid s_{t+1}] &= \exp(-r_t - a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} - b(\tau-1)'G^{\mathbb{Q}}f_t) \\ &\quad \times \exp\left(\frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right)\end{aligned}$$

$$\begin{aligned}
&= \exp\left(-a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} + \frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right) \\
&\quad \times \exp\left(-\left(\beta' + b(\tau-1)'G^{\mathbb{Q}}\right)f_t\right),
\end{aligned}$$

where the last equality used the fact that $r_t = \delta_{s_t} + \beta'f_t$. Therefore,

$$\begin{aligned}
\exp\left(-a_{s_t}(\tau) - b(\tau)'f_t\right) &= \mathbb{E}_t^{\mathbb{Q}}\left[\mathbb{E}_t^{\mathbb{Q}}\left[\exp(-r_t)P_{t+1}(\tau-1) \mid s_{t+1}\right]\right] \\
&= \exp\left(-\left(\beta' + b(\tau-1)'G^{\mathbb{Q}}\right)f_t\right) \\
&\quad \times \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} + \frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right)\right].
\end{aligned}$$

Taking logs on both sides, we obtain

$$\begin{aligned}
-a_{s_t}(\tau) - b(\tau)'f_t &= -\left(\beta' + b(\tau-1)'G^{\mathbb{Q}}\right)f_t \\
&\quad + \log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} + \frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right)\right]\right),
\end{aligned}$$

and matching the intercept and coefficient terms on both sides yields the recursions

$$\begin{aligned}
a_{s_t}(\tau) &= -\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-a_{s_{t+1}}(\tau-1) - b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} + \frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right)\right]\right) \\
b(\tau) &= \beta + G^{\mathbb{Q}}b(\tau-1)
\end{aligned}$$

for any $\tau \geq 1$. If we ignore Jensen's inequality to interchange the log and expectation, we have the simpler expression

$$a_{s_t}(\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[a_{s_{t+1}}(\tau-1) + b(\tau-1)'K_{s_{t+1}}^{\mathbb{Q}} - \frac{1}{2}b(\tau-1)'\Omega_{s_{t+1}}b(\tau-1)\right].$$

These recursions, together with the initial conditions $a_{s_t}(0) = 0$ and $b(0) = O_{n \times 1}$, allow us to obtain closed form solutions for the functions $a_{s_t}(\cdot)$ and $b(\cdot)$.

These closed form solutions are given as follows: starting with the easier function,

it is clear that $b(1) = \beta$. Suppose that $b(\tau) = \sum_{i=0}^{\tau-1} (G^{\mathbb{Q}'})^i \beta$ for some maturity $\tau > 1$. Then, by induction,

$$b(\tau+1) = \beta + G^{\mathbb{Q}'} b(\tau) = \beta + \sum_{i=1}^{\tau} (G^{\mathbb{Q}'})^i \beta = \sum_{i=0}^{\tau} (G^{\mathbb{Q}'})^i \beta.$$

It follows by induction that, for any maturity τ ,

$$b(\tau) = \sum_{i=0}^{\tau-1} (G^{\mathbb{Q}'})^i \beta.$$

Moving onto $a_{s_t}(\cdot)$, note that $a_{s_t}(1) = 0$. Under the first order approximation that allows us to ignore Jensen's inequality, we have

$$a_{s_t}(\tau) = \mathbb{E}_t^{\mathbb{Q}} [a_{s_{t+1}}(\tau-1)] + \mathbb{E}_t^{\mathbb{Q}} \left[b(\tau-1)' K_{s_{t+1}}^{\mathbb{Q}} - \frac{1}{2} b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right].$$

It follows from an induction process, similar to the one previously employed, that

$$a_{s_t}(\tau) = \sum_{i=1}^{\tau-1} b(\tau-i)' \mathbb{E}_t^{\mathbb{Q}} [K_{s_{t+i}}^{\mathbb{Q}}] - \frac{1}{2} \sum_{i=1}^{\tau-1} b(\tau-i)' \mathbb{E}_t^{\mathbb{Q}} [\Omega_{s_{t+i}}] b(\tau-i).$$

The Markov property of the regime process tells us that the $\mathbb{Q}$ -expectations above can be expanded as

$$a_{s_t}(\tau) = \sum_{i=1}^{\tau-1} b(\tau-i)' \sum_{j=1}^2 P_{s_t,j}^{(i)} \cdot K_j^{\mathbb{Q}} - \frac{1}{2} \sum_{i=1}^{\tau-1} b(\tau-i)' \left(\sum_{j=1}^2 P_{s_t,j}^{(i)} \Omega_j \right) b(\tau-i),$$

where $P_{ij}^{(i)}$ is the (i,j) th element of the matrix product P^i . In other words,

$$a_{s_t}(\tau) = -\frac{1}{2} \sum_{i=1}^{\tau-1} b(\tau-i)' \left(\sum_{j=1}^2 P_{s_t,j}^{(i)} \Omega_j \right) b(\tau-i) + \sum_{j=1}^2 \left(\sum_{i=1}^{\tau-1} b(\tau-i)' P_{s_t,j}^{(i)} \right) K_j^{\mathbb{Q}}, \quad (\text{A.1})$$

which tells us that $a_{s_t}(\tau)$ is affine in the vector $(K_1^{\mathbb{Q}}, K_2^{\mathbb{Q}})$.

B Derivation of Modified AFNS Factor Loadings

Let L_t^*, S_t^* and C_t^* be the factors from the original AFNS model, and L_t, S_t , and C_t the factors from the modified model. The original AFNS model implies that follows that

$$Y_t(\tau) = \frac{a_{s_t}(\tau)}{\tau} + L_t^* + \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} S_t^* + \left(\frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} - \exp(-\tau\kappa^{\mathbb{Q}}) \right) C_t^*.$$

Defining $L_t = L_t^* + S_t^*$, $S_t = -S_t^*$ and $C_t = C_t^*$, we can rearrange the terms of the above equation to see that

$$\begin{aligned} Y_t(\tau) &= \frac{a_{s_t}(\tau)}{\tau} + L_t + S_t - \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} S_t + \left(\frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} - \exp(-\tau\kappa^{\mathbb{Q}}) \right) C_t \\ &= \frac{a_{s_t}(\tau)}{\tau} + L_t + \left(1 - \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} \right) S_t + \left(\frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} - \exp(-\tau\kappa^{\mathbb{Q}}) \right) C_t. \end{aligned}$$

Labeling the new factors L_t , S_t , C_t as the level, slope and curvature factors respectively, the factor loading on the level factor is constant, while the loading on the curvature factor is unchanged from the original AFNS model. Meanwhile, the loading on the slope factor begins at 0 and increases to 1 as τ goes from 0 to ∞ . Because the slope and curvature factors have near-zero loadings for short maturities, the level factor now represents the short end of the yield curve, that is,

$$r_t = \beta' f_t \approx L_t.^1$$

¹This equation is a key reason for choosing the AFNS model over the specification introduced by Joslin, Singleton, and Zhu (2011) (JSZ). In the modified AFNS model, the level factor is clearly identified as the factor that represents the short end of the yield curve, but interpreting the level factor in this manner under the JSZ specification proves much more difficult.

Furthermore, while $G^{\mathbb{Q}}$ in the JSZ model may depend on three distinct eigenvalues, in the AFNS specification it relies on just a single parameter $\kappa^{\mathbb{Q}}$. This results in a more parsimonious model, which

The important part is that the modified model is simply the original AFNS model with rearranged factors, and as such, that its capability to fit the yield curve is identical to that of the original model. This is confirmed in the section on our model results, which shows our model's ability to fit the yield curve with minimal measurement errors.

The loadings of the factors on the yields are now determined as

$$\frac{b(\tau)}{\tau} = \left(1 \quad 1 - \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} \quad \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\tau\kappa^{\mathbb{Q}}} - \exp(-\tau\kappa^{\mathbb{Q}}) \right)'.$$

We now derive expressions for β and $G^{\mathbb{Q}}$ that yield these factor loadings under the pricing formula obtained above.

Note first that

$$b(\tau)' = \beta' + b(\tau - 1)'^{\mathbb{Q}}$$

for any $\tau \geq 1$, where $b(0) = O_{n \times 1}$. Suppose β and $G^{\mathbb{Q}}$ are given as

$$\beta = \left(1 \quad 1 - \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} \quad \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \exp(-\kappa^{\mathbb{Q}}) \right)' \quad (\text{B.1})$$

$$G^{\mathbb{Q}} = \begin{pmatrix} 1 & 1 - \exp(-\kappa^{\mathbb{Q}}) & \kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ 0 & \exp(-\kappa^{\mathbb{Q}}) & -\kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ 0 & 0 & \exp(-\kappa^{\mathbb{Q}}) \end{pmatrix}. \quad (\text{B.2})$$

Then,

$$\beta' = b(1)' = \left(1 \quad 1 - \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} \quad \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \exp(-\kappa^{\mathbb{Q}}) \right).$$

is an additional reason for preferring the AFNS model to the JSZ alternative.

Now suppose that $\frac{b(\tau)}{\tau}$ collects the modified AFNS factor loadings for any $\tau \geq 1$. Then,

$$\begin{aligned}
\beta' + b(\tau)'^{\mathbb{Q}} &= \begin{pmatrix} 1 & 1 - \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \exp(-\kappa^{\mathbb{Q}}) \end{pmatrix} \\
&\quad + \begin{pmatrix} \tau & \tau - \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \frac{1 - \exp(-\tau\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \tau \exp(-\tau\kappa^{\mathbb{Q}}) \end{pmatrix} \begin{pmatrix} 1 & 1 - \exp(-\kappa^{\mathbb{Q}}) & \kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ 0 & \exp(-\kappa^{\mathbb{Q}}) & -\kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ 0 & 0 & \exp(-\kappa^{\mathbb{Q}}) \end{pmatrix} \\
&= \begin{pmatrix} \tau & \tau - \frac{\exp(-\kappa^{\mathbb{Q}}) - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \exp(-\kappa^{\mathbb{Q}}) - (\tau+1)\exp(-(\tau+1)\kappa^{\mathbb{Q}}) + \frac{\exp(-\kappa^{\mathbb{Q}}) - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} \end{pmatrix} \\
&\quad + \begin{pmatrix} 1 & 1 - \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \frac{1 - \exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \exp(-\kappa^{\mathbb{Q}}) \end{pmatrix} \\
&= \begin{pmatrix} \tau+1 & \tau+1 - \frac{1 - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \frac{1 - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - (\tau+1)\exp(-(\tau+1)\kappa^{\mathbb{Q}}) \end{pmatrix} = b(\tau+1)',
\end{aligned}$$

so that

$$\frac{b(\tau+1)}{\tau+1} = \begin{pmatrix} 1 & 1 - \frac{1 - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{(\tau+1)\kappa^{\mathbb{Q}}} & \frac{1 - \exp(-(\tau+1)\kappa^{\mathbb{Q}})}{(\tau+1)\kappa^{\mathbb{Q}}} - \exp(-(\tau+1)\kappa^{\mathbb{Q}}) \end{pmatrix}',$$

so that, by induction, $\frac{b(\tau)}{\tau}$ are exactly the modified AFNS factor loadings for any $\tau \geq 1$.

C Determination of Physical Factor Dynamics

C.1 Incorporating Falling Stars into the AFNS Model

Here, we momentarily omit the regime subscript for notational simplicity. Campbell and Shiller (1987) note that yields are best modeled as cointegrated with a single common trend. To reflect this finding, Bauer and Rudebusch (2020) specify the $\mathbb{P}$ -dynamics of the factors as a trend-cycle VAR(1) of the form

$$f_t = \gamma \cdot \tau_t + \tilde{f}_t,$$

where γ represents the loading of the common trend τ_t on the factors, and $\tilde{f}_t = (l_t, S_t, C_t)'$ follows a stationary VAR(1). Because there is only one common trend τ_t , there are $3 - 1 = 2$ cointegrating relationships among the factors. Yields are affine functions of f_t , so this specification ensures that yields are cointegrated with exactly one common trend.

Non-stationary AFNS models such as those introduced by Kopp and P. D. Williams (2018), Cao et al. (2020), and Brand, Goy, and Lemke (2021) impose the additional assumption that $\gamma = (1, 0, 0)'$. This indicates that the slope and curvature factors are jointly stationary, while the common trend loads on the level factor with unity. Under this specification, yield spreads are stationary. This is in accordance with the findings in Hall, Anderson, and Granger (1992), which are further corroborated by Brand, Goy, and Lemke (2021).

When estimating models whose $\mathbb{P}$ -dynamics are specified as above, the state-space representation of the model includes four unobserved components— τ_t, l_t, S_t , and C_t — and the trend and cycle components of the level factor are estimated separately. Unlike previous AFNS models with falling stars, the focus of our model is on the incorporation of the ZLB into an ATSM with falling stars, rather than on the estimation of macroeconomic stars. Therefore, the $\mathbb{P}$ -dynamics of the factors should ideally be presented in a way that maintains the cointegrating relationships among the yields, but does not include τ_t , a variable that is difficult to estimate and of secondary importance in this model.

To this end, we use the following result.

Proposition 1. *Suppose $f_t = (L_t, S_t, C_t)'$ follows the VAR(1) process*

$$f_t = G^{\mathbb{P}} f_{t-1} + \Omega^{\frac{1}{2}} \cdot v_t^{\mathbb{P}} \tag{C.1}$$

$$v_t^{\mathbb{P}} \mid \mathcal{F}_{t-1}, s_t \sim \mathcal{N}(O_{3 \times 1}, I_3) \text{ under } \mathbb{P}.$$

The level factor is $I(1)$ and the slope and curvature factors are jointly $I(0)$ if and only if

$$G^{\mathbb{P}} = \begin{pmatrix} 1 & G_{12}^{\mathbb{P}} & G_{13}^{\mathbb{P}} \\ 0 & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ 0 & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix}, \quad \max \left| \text{eig} \begin{pmatrix} G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix} \right| < 1, \quad \text{and} \quad \omega^2 > 0, \quad (\text{C.2})$$

where ω^2 is the innovation variance of the level factor, and the operation eig recovers the eigenvalues of a given square matrix ².

Proof. See Appendix D. □

The VAR(1) representation above can thus replicate the cointegration properties of the factors found in the usual trend-cycle representation, without having to add a separate trend component. This leads to a more parsimonious model, and in terms of estimation, it eliminates the pile-up problem, since it is no longer necessary to identify the trend and cycle components of the factors separately.

C.2 Incorporating the ZLB through Regime Shifts

We now include the ZLB into the falling stars framework described earlier. Although our approach differs from shadow rate models in that we do not explicitly require the short rate to be non-negative, we can ensure that the likelihood of the short rate turning negative is extremely low by assuming that the short rate oscillates around 0 without much volatility when it comes sufficiently close to 0.

Formally, we consider the economy to be in one of two exogenously determined regimes: the normal regime ($s_t = 1$) and the ZLB regime ($s_t = 2$). In the normal regime,

²The conditions presented here are almost identical to those found in the appendix of Bauer and Rudebusch (2020). There, the conditions in (C.2) are recovered as necessary conditions of the trend-cycle decomposition. By demonstrating that these conditions are also sufficient for the cointegration properties of the factors, we establish that the VAR and trend-cycle representations are equivalent.

the data generating process adheres to the model described in the previous section, meaning that the level factor is $I(1)$ and the other two factors are jointly $I(0)$. The AFNS factor loadings then imply that any m -dimensional vector of bond yields is cointegrated with cointegration rank $m - 1$ and exactly one common stochastic trend. The mean reversion coefficient during the normal regime under the physical measure is given according to (C.2)

$$G_1^{\mathbb{P}} = \begin{pmatrix} 1 & G_{12,1}^{\mathbb{P}} & G_{13,1}^{\mathbb{P}} \\ 0 & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ 0 & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix},$$

where the subscript 1 indicates that the formulation holds for the normal regime.

At the lower bound, the short rate exhibits minimal volatility and is stationary around 0. This means that the mean reversion parameter during the lower bound regime must be given by

$$G_2^{\mathbb{P}} = \begin{pmatrix} \rho_2 & 0 & 0 \\ 0 & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ 0 & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix},$$

where $|\rho_2| < 1$, and that the variance ω_2^2 of the level factor is strictly smaller than its regime 1 counterpart, ω_1^2 ³.

By allowing the coefficient on L_{t-1} to regime shift from 1 to ρ_2 , a value that is less than unity in magnitude, we are essentially assuming that, during the lower bound

³In Hördahl and Tristani (2019), where the economy switches between a normal and lower bound regime similar to our paper, the short rate is assumed to be a white noise process oscillating around a constant during the lower bound regime. While conceptually simple, we find that there are some discrepancies between the white noise specification and the data. For instance, the OLS estimate of ρ from the regression $r_t = \rho r_{t-1} + u_t$ for short rate values between January 2009 and December 2015 is given as $\hat{\rho} = 0.9859$, a value very close to 1. This suggests that, far from being a white noise process, the short rate is better modeled as a persistent yet stationary process during the lower bound regime.

regime, the stochastic common trend τ_t in Equation (C.1) that drives the factors is fixed at 0. In other words, the stochastic trend disappears during the lower bound regime, reflecting the fact that short term interest rates do not exhibit large downward movements when the economy is at the lower bound.

D Proof of Proposition 1

Note first that the factor innovation covariance matrix Ω can be decomposed as

$$\Omega = \begin{pmatrix} \omega^2 & O_{1 \times 2} \\ O_{2 \times 1} & \Omega_{22} \end{pmatrix}.$$

Naturally, its Cholesky factor is

$$\Omega^{\frac{1}{2}} = \begin{pmatrix} \omega & O_{1 \times 2} \\ O_{2 \times 1} & \Omega_{22}^{\frac{1}{2}} \end{pmatrix}.$$

D.1 Sufficiency

We first show sufficiency. Suppose $G^{\mathbb{P}}$ and Ω satisfy the conditions laid out in proposition

1. The VECM representation of f_t is given as

$$\Delta f_t = \underbrace{(G^{\mathbb{P}} - I_3)}_{\Pi} f_{t-1} + \Omega^{\frac{1}{2}} v_t^{\mathbb{P}}, \quad (\text{D.1})$$

where

$$\Pi = G^{\mathbb{P}} - I_3 = \begin{pmatrix} O_{n \times 1} & \nu_1 & \nu_2 \end{pmatrix}$$

for 3-dimensional vectors ν_1 and ν_2 . Since ν_1 and ν_2 are

$$\begin{pmatrix} \nu_1 & \nu_2 \end{pmatrix} = \begin{pmatrix} G_{12}^{\mathbb{P}} & G_{13}^{\mathbb{P}} \\ G_{22}^{\mathbb{P}} - 1 & G_{23}^{\mathbb{P}} \\ G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} - 1 \end{pmatrix}$$

and $(G_{22}^{\mathbb{P}} - 1)(G_{33}^{\mathbb{P}} - 1) - G_{23}^{\mathbb{P}}G_{32}^{\mathbb{P}} \neq 0$ by the eigenvalue condition, the lower 2×2 block in the matrix above has a non-zero determinant. This implies that ν_1 and ν_2 are linearly independent. Note that Π can be decomposed as

$$\Pi = \begin{pmatrix} \nu_1 & \nu_2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = AB',$$

where A is a full rank 3×2 matrix with columns ν_1 and ν_2 .

Define $A_{\perp}$ and $B_{\perp}$ as orthogonal complements of A and B , that is, non-zero 3-dimensional vectors such that $B'B_{\perp} = O_{2 \times 1}$ and $A'A_{\perp} = O_{2 \times 1}$; $\{A_{\perp}\}$ can be interpreted as a basis of the orthogonal complement of the vector space spanned by the columns of A , and likewise for $B_{\perp}$. In light of the latter interpretation, we can take $B_{\perp} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}'$.

We can also show that the first element of $A_{\perp}$ must be non-zero. Suppose that $A_{\perp}$ is given as

$$A_{\perp} = \begin{pmatrix} 0 & a & b \end{pmatrix}'$$

for some $a, b \in \mathbb{R}$. Then,

$$A'_\perp A = \begin{pmatrix} 0 & a & b \end{pmatrix} \begin{pmatrix} G_{12}^\mathbb{P} & G_{13}^\mathbb{P} \\ G_{22}^\mathbb{P} - 1 & G_{23}^\mathbb{P} \\ G_{32}^\mathbb{P} & G_{33}^\mathbb{P} - 1 \end{pmatrix} = \begin{pmatrix} a(G_{22}^\mathbb{P} - 1) + bG_{32}^\mathbb{P} & aG_{23}^\mathbb{P} + b(G_{33}^\mathbb{P} - 1) \end{pmatrix} = O_{1 \times 2}.$$

Because $A_\perp$ is non-zero, either a or b is non-zero. Without loss of generality, suppose that $a \neq 0$. Then,

$$\begin{aligned} G_{22}^\mathbb{P} - 1 &= -\frac{b}{a}G_{32}^\mathbb{P} \\ G_{23}^\mathbb{P} &= -\frac{b}{a}(G_{33}^\mathbb{P} - 1), \end{aligned}$$

so that

$$(G_{22}^\mathbb{P} - 1)(G_{33}^\mathbb{P} - 1) - G_{23}^\mathbb{P}G_{32}^\mathbb{P} = 0,$$

which contradicts our eigenvalue assumption. Therefore, the first element of $A_\perp$ cannot be 0, and we normalize it to 1.

So far, we have seen that $\Pi = AB'$ for 3×2 matrices A, B of full rank, and we can choose orthogonal complements $A_\perp$ and $B_\perp$ of A and B so that $A'_\perp B_\perp = \mathbf{1} \neq 0$. Therefore, the conditions for the Granger representation theorem to hold are all satisfied (Theorem 4.1 in Johansen (1991)). By the Granger representation theorem, f_t has the Beveridge-Nelson decomposition

$$f_t = \Psi \cdot \left(\sum_{s=1}^t \varepsilon_s^\mathbb{P} \right) + \eta_t - \eta_0 + f_0$$

for any $1 \leq t \leq T$, where η_t is the stationary part of f_t and $\sum_{s=1}^t \varepsilon_s^\mathbb{P}$ is the trend part of

f_t . Ψ is given as

$$\Psi = B_{\perp} \left(A'_{\perp} B_{\perp} \right)^{-1} A'_{\perp},$$

where

$$B_{\perp} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}',$$

and $A'_{\perp} B_{\perp} = 1$, so that

$$\Psi = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} A'_{\perp}.$$

Defining the single common stochastic trend τ_t as

$$\tau_t = \sum_{s=1}^t A'_{\perp} \varepsilon_s^{\mathbb{P}},$$

the common factors f_t have the common trend representation

$$f_t = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tau_t + \eta_t,$$

provided that $\eta_0 = f_0$. This is exactly the common trend representation of the factors used in previous AFNS models in the falling stars literature.

The 0 loadings of the trend on the slope and curvature factors reveals that they are

jointly stationary. Moreover, the first equation in the physical factor dynamics is

$$L_t = L_{t-1} + G_{12}^{\mathbb{P}} S_{t-1} + G_{13}^{\mathbb{P}} C_{t-1} + \varepsilon_{1,t}^{\mathbb{P}},$$

where $\varepsilon_{1,t}^{\mathbb{P}}$ has positive variance $\omega^2 > 0$, so that the level factor is indeed non-stationary.

The first difference of the level factor is the stationary process

$$\Delta L_t = G_{12}^{\mathbb{P}} S_{t-1} + G_{13}^{\mathbb{P}} C_{t-1} + \varepsilon_{1,t}^{\mathbb{P}},$$

meaning that L_t is an $I(1)$ variable whose first difference ΔL_t is a stationary process whose dynamics are governed by the dynamics of the slope and curvature factors.

Therefore, if $G^{\mathbb{P}}$ and Ω satisfy the conditions of proposition 1, then the level factor is $I(1)$ while the slope and curvature factors are jointly $I(0)$.

D.2 Necessity

Conversely, suppose that L_t is $I(1)$ and S_t, C_t are jointly $I(0)$. Note that the first equation in the error correction representation (D.1) is

$$\Delta L_t = \Pi_{11} L_{t-1} + \Pi_{12} S_{t-1} + \Pi_{13} C_{t-1} + \varepsilon_{1,t}^{\mathbb{P}}$$

Because ΔL_t is stationary and L_{t-1} is non-stationary, we must have $\Pi_{11} = 0$.

In addition, by assumption, $B' f_t$ is a 2-dimensional stationary time series, where $B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. From the VAR representation of the factors we have

$$B' f_t = B' G^{\mathbb{P}} f_{t-1} + B' \Omega^{\frac{1}{2}} \cdot v_t^{\mathbb{P}}.$$

L_{t-1} is non-stationary and

$$B'G^{\mathbb{P}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} G_{11}^{\mathbb{P}} & G_{12}^{\mathbb{P}} & G_{13}^{\mathbb{P}} \\ G_{21}^{\mathbb{P}} & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ G_{31}^{\mathbb{P}} & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix} = \begin{pmatrix} G_{21}^{\mathbb{P}} & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ G_{31}^{\mathbb{P}} & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix},$$

so we must have $G_{21}^{\mathbb{P}} = G_{31}^{\mathbb{P}} = 0$. Since

$$\Pi = G^{\mathbb{P}} - I_3 = \begin{pmatrix} G_{11}^{\mathbb{P}} - 1 & G_{12}^{\mathbb{P}} & G_{13}^{\mathbb{P}} \\ G_{21}^{\mathbb{P}} & G_{22}^{\mathbb{P}} - 1 & G_{23}^{\mathbb{P}} \\ G_{31}^{\mathbb{P}} & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} - 1 \end{pmatrix},$$

this implies that $\Pi_{21} = \Pi_{31} = 0$. As such, the first column of Π must be equal to 0.

The above finding indicates that Π must be a reduced rank matrix, or that its rank must be 0, 1 or 2. If the rank of Π is 0, then $\Pi = O_{3 \times 3}$ and $G^{\mathbb{P}} = I_3$; this would mean that all three factors follow a random walk and thus that all three factors, not just the level factor, are non-stationary. This contradicts the assumed stationarity of the slope and curvature factors, so the rank of Π cannot be 0.

Suppose the rank of Π is 1. Then, its columns must be scalar multiples of one another. Without loss of generality, suppose there exists a $c \in \mathbb{R}$ such that

$$c \begin{pmatrix} G_{12}^{\mathbb{P}} \\ G_{22}^{\mathbb{P}} - 1 \\ G_{32}^{\mathbb{P}} \end{pmatrix} = \begin{pmatrix} G_{13}^{\mathbb{P}} \\ G_{23}^{\mathbb{P}} \\ G_{33}^{\mathbb{P}} - 1 \end{pmatrix}. \quad (\text{D.2})$$

Noting that $G_{21}^{\mathbb{P}} = G_{31}^{\mathbb{P}} = 0$, the last two equations comprising the VAR representation

of the factors form their own 2-dimensional VAR(1) system given as

$$\begin{pmatrix} S_t \\ C_t \end{pmatrix} = \begin{pmatrix} G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ C_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{2,t}^{\mathbb{P}} \\ \varepsilon_{3,t}^{\mathbb{P}} \end{pmatrix}.$$

Because (S_t, C_t) are jointly stationary, the eigenvalues of

$$\begin{pmatrix} G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix} \tag{D.3}$$

must not lie on or outside the unit circle, which implies that the matrix

$$\begin{pmatrix} G_{22}^{\mathbb{P}} - 1 & G_{23}^{\mathbb{P}} \\ G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} - 1 \end{pmatrix}$$

must have a non-zero determinant (otherwise, 1 is an eigenvalue of the matrix in (D.3)).

However, if equation (D.2) holds, then

$$(G_{22}^{\mathbb{P}} - 1)(G_{33}^{\mathbb{P}} - 1) - G_{23}^{\mathbb{P}}G_{32}^{\mathbb{P}} = c \cdot (G_{22}^{\mathbb{P}} - 1)G_{32}^{\mathbb{P}} - c \cdot (G_{22}^{\mathbb{P}} - 1)G_{32}^{\mathbb{P}} = 0,$$

a contradiction. Therefore, the rank of Π cannot be 1.

It follows that the rank of Π must be 2, and as such, Π must assume the form

$$\Pi = \begin{pmatrix} O_{3 \times 1} & \nu_1 & \nu_2 \end{pmatrix}$$

for linearly independent 3-dimensional vectors ν_1 and ν_2 . This implies that $G^{\mathbb{P}}$ is given

as

$$G^{\mathbb{P}} = \Pi + I_3 = \begin{pmatrix} 1 & G_{12}^{\mathbb{P}} & G_{13}^{\mathbb{P}} \\ 0 & G_{22}^{\mathbb{P}} & G_{23}^{\mathbb{P}} \\ 0 & G_{32}^{\mathbb{P}} & G_{33}^{\mathbb{P}} \end{pmatrix}.$$

We showed above that matrix (D.3) must have eigenvalues that lie within the unit circle in order for the joint stationarity of S_t, C_t to hold.

Finally, under the above specification

$$L_t = L_{t-1} + G_{12}^{\mathbb{P}} S_{t-1} + G_{13}^{\mathbb{P}} C_{t-1} + \omega \cdot v_{1,t}^{\mathbb{P}},$$

so in order for L_t to be non-stationary, $\omega > 0$. □

E Market Prices of Risk and Excess Returns

Here we fill in the details about the choice of SDF and the market prices of risk, as well as the formula for the one-period excess bond returns that is used to derive the decomposition for the term premium. In this section we state the relevant results, and the proofs are left for subsequent Appendices.

E.1 The SDF and Market Prices of Risk

As stated in the main text, the stochastic discount factor (SDF) is given in the usual stochastic exponential form:

$$\mathcal{M}_{t+1,s_{t+1}} = \exp\left(-r_t - \frac{1}{2} \lambda'_{t,s_{t+1}} \lambda_{t,s_{t+1}} - \lambda'_{t,s_{t+1}} v_{t+1}^{\mathbb{P}}\right),$$

where $\lambda_{t,s_{t+1}}$ is the market price of risk, which here is dependent on the one-period ahead regime s_{t+1} . The SDF enables us to move from the risk-neutral measure to the physical measure, and vice versa. Specifically, for any variable X_{t+1} known from time $t+1$ onward, the following relationship holds:

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}[X_{t+1}] &= \mathbb{E}_t \left[X_{t+1} \exp(r_t) \mathcal{M}_{t+1,s_{t+1}} \right] \\ &= \mathbb{E}_t \left[X_{t+1} \exp \left(-\frac{1}{2} \lambda'_{t,s_{t+1}} \lambda_{t,s_{t+1}} - \lambda'_{t,s_{t+1}} v_{t+1}^{\mathbb{P}} \right) \right].\end{aligned}$$

As shown in Appendix F, letting

$$v_{t+1}^{\mathbb{Q}} = v_{t+1}^{\mathbb{P}} + \lambda_{t,s_{t+1}}, \quad (\text{E.1})$$

$v_{t+1}^{\mathbb{Q}} \mid \mathcal{F}_{t,s_{t+1}}$ follows the normal distribution $\mathcal{N}(O_{3 \times 1}, I_3)$ under the risk-neutral measure $\mathbb{Q}$ if the same holds for $v_{t+1}^{\mathbb{P}} \mid \mathcal{F}_{t,s_{t+1}}$ under the physical measure.

We opt for the essentially affine representation of the market prices of risk, as in Cheridito, Filipovic, and Kimmel (2007), implied by the risk-neutral and physical dynamics. In Appendix G, we show that $\lambda_{t,s_{t+1}}$ is given as

$$\begin{aligned}\lambda_{t,s_{t+1}} &= \lambda_{s_{t+1}} + \Lambda_{s_{t+1}} \cdot f_t, \\ \lambda_{s_{t+1}} &= - \left(\Omega_{s_{t+1}}^{\frac{1}{2}} \right)^{-1} K_{s_{t+1}}^{\mathbb{Q}} \\ \Lambda_{s_{t+1}} &= \left(\Omega_{s_{t+1}}^{\frac{1}{2}} \right)^{-1} \left(G_{s_{t+1}}^{\mathbb{P}} - G_{s_{t+1}}^{\mathbb{Q}} \right).\end{aligned} \quad (\text{E.2})$$

E.2 Excess Returns and Risk premia

The one-period ahead holding period return of a τ -maturity bond at time t is

$$hpr_{t+1}^{(\tau)} = \log(P_{t+1}(\tau - 1)) - \log(P_t(\tau)).$$

The expected excess return to a τ -maturity bond from time t to $t+1$ is then defined as

$$exr_{t+1}^{(\tau)} = \mathbb{E}_t \left[hpr_{t+1}^{(\tau)} \right] - r_t.$$

When we ignore Jensen's inequality, so that logs and expectations may be interchanged, the expected excess return formula is given as

$$\begin{aligned} exr_{t+1}^{(\tau)} &= b(\tau-1)' \mathbb{E}_t \left[K_{s_{t+1}}^{\mathbb{Q}} \right] - \frac{1}{2} b(\tau-1)' \mathbb{E}_t \left[\Omega_{s_{t+1}} \right] b(\tau-1) \\ &\quad - b(\tau-1)' \mathbb{E}_t \left[\left(G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}} \right) f_t \right], \end{aligned}$$

where the expectations are written in a way that makes explicit that the regime at time t is known. This expression is also derived in Appendix G.

F The Discrete Version of Girsanov's Theorem

Let $v_{t+1}^{\mathbb{Q}} = v_{t+1}^{\mathbb{P}} + \lambda_{t,s_{t+1}}$, and suppose the SDF $\mathcal{M}_{t+1,s_{t+1}}$ is given as

$$\mathcal{M}_{t+1,s_{t+1}} = \exp \left(-r_t - \frac{1}{2} \lambda'_{t,s_{t+1}} \lambda_{t,s_{t+1}} - \lambda'_{t,s_{t+1}} v_{t+1}^{\mathbb{P}} \right).$$

We want to show that

$$\mathbb{E}_t^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \mid s_{t+1} \right] = \exp \left(\frac{1}{2} r' r \right)$$

for any $r \in \mathbb{R}^3$. Then, because the moment generating function of a random vector determines its distribution, and the right hand side is precisely the moment generating function of a standard normally distributed random vector, we can conclude that

$$v_{t+1}^{\mathbb{Q}} \mid \mathcal{F}_t, s_{t+1} \sim \mathcal{N}(O_{3 \times 1}, I_3).$$

We first prove three smaller results before moving onto Girsanov's theorem. Below, $\sigma(s_{t+1})$ denotes the σ -algebra generated by s_{t+1} , and

$$\mathcal{G}_t = \sigma(s_{t+1}) \vee \mathcal{F}_t$$

the σ -algebra generated by the union $\sigma(s_{t+1}) \cup \mathcal{F}_t$.

Lemma F.1. *For any $r \in \mathbb{R}^3$ and $B \in \sigma(s_{t+1})$,*

$$\mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \cdot I_B \right] = \mathbb{E}_t^{\mathbb{Q}} \left[\exp\left(\frac{1}{2} r' r\right) \cdot I_B \right].$$

Proof. The following chain of equalities hold:

$$\begin{aligned} \mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \cdot I_B \right] &= \mathbb{E}_t \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \mathcal{M}_{t+1, s_{t+1}} \exp(r_t) \cdot I_B \right] \\ &= \mathbb{E}_t \left[\exp\left(-r' \lambda_{t, s_{t+1}} - \frac{1}{2} \lambda'_{t, s_{t+1}} \lambda_{t, s_{t+1}} - (r + \lambda_{t, s_{t+1}})' v_{t+1}^{\mathbb{P}}\right) \cdot I_B \right] \\ &= \mathbb{E}_t \left[\exp\left(-r' \lambda_{t, s_{t+1}} - \frac{1}{2} \lambda'_{t, s_{t+1}} \lambda_{t, s_{t+1}}\right) I_B \cdot \mathbb{E}_t \left[\exp(-(r + \lambda_{t, s_{t+1}})' v_{t+1}^{\mathbb{P}}) \mid s_{t+1} \right] \right], \end{aligned}$$

where the first equality is due to the definition of the risk-neutral measure, the second due to our specification of the SDF, and the third due to the law of iterated expectations. Since $v_{t+1}^{\mathbb{P}}$ follows a standard normal distribution given $\mathcal{F}_t$ and s_{t+1} , we have

$$\begin{aligned} \mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \cdot I_B \right] &= \mathbb{E}_t \left[\exp\left(\frac{1}{2} r' r\right) \cdot I_B \right] \\ &= \exp\left(\frac{1}{2} r' r\right) \mathbb{E}_t [I_B]. \end{aligned}$$

By the Doob-Dynkin Lemma, that $B \in \sigma(s_{t+1})$ tells us that there exists a (non-negative) function $f : \{1, \dots, N\} \rightarrow \mathbb{R}$ such that $I_B = f(s_{t+1})$; this function is trivially measurable

because its domain is finite. The expectation above can now be written as

$$\mathbb{E}_t[I_B] = \mathbb{E}[f(s_{t+1}) \mid s_t] = \sum_{i=1}^N f(i) \cdot \mathbb{P}(s_{t+1} = i \mid s_t).$$

where the first equality follows from the Markov property of the regime process. Since

$$\mathbb{P}(s_{t+1} = i \mid s_t) = \mathbb{Q}(s_{t+1} = i \mid s_t) = P_{s_t, i}$$

by assumption, it follows that

$$\mathbb{E}_t[I_B] = \sum_{i=1}^N f(i) \cdot \mathbb{Q}[s_{t+1} = i \mid s_t] = \mathbb{E}^{\mathbb{Q}}[f(s_{t+1}) \mid s_t] = \mathbb{E}_t^{\mathbb{Q}}[f(s_{t+1})] = \mathbb{E}_t^{\mathbb{Q}}[I_B].$$

As such,

$$\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r'v_{t+1}^{\mathbb{Q}}\right) \cdot I_B\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(\frac{1}{2}r'r\right) \cdot I_B\right].$$

□

Lemma F.2. *For any $r \in \mathbb{R}^3$, $A \in \mathcal{F}_t$ and $B \in \sigma(s_{t+1})$, we have*

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r'v_{t+1}^{\mathbb{Q}}\right) \mid s_{t+1}\right] \cdot I_{A \cap B}\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\frac{1}{2}r'r\right) \cdot I_{A \cap B}\right].$$

Proof. Here we use the notation $\mathbb{E}^{\mathbb{Q}}[\cdot \mid \mathcal{G}_t]$ and $\mathbb{E}^{\mathbb{Q}}[\cdot \mid \mathcal{F}_t]$ for $\mathbb{E}_t^{\mathbb{Q}}[\cdot \mid s_{t+1}]$ and $\mathbb{E}_t^{\mathbb{Q}}[\cdot]$ to emphasize the conditioning sets for the respective conditional expectations.

Choose any $A \in \mathcal{F}_t$ and $B \in \sigma(s_{t+1})$. $A \cap B \in \mathcal{G}_t$ because σ -algebras are closed under finite intersections, so that

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r'v_{t+1}^{\mathbb{Q}}\right) \mid \mathcal{G}_t\right] \cdot I_{A \cap B}\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-r'v_{t+1}^{\mathbb{Q}}\right) \cdot I_{A \cap B}\right]$$

from the definition of conditional expectations. Now, we can see that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \cdot I_{A \cap B} \right] &= \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \cdot I_B \mid \mathcal{F}_t \right] \cdot I_A \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{1}{2} r' r \right) \cdot I_B \mid \mathcal{F}_t \right] \cdot I_A \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{1}{2} r' r \right) \cdot I_{A \cap B} \mid \mathcal{F}_t \right] \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{1}{2} r' r \right) \cdot I_{A \cap B} \right],
\end{aligned}$$

where the first equality follows from the definition of conditional expectations and the $\mathcal{F}_t$ -measurability of A , the second from lemma D.1, the third from the fact that I_A is $\mathcal{F}_t$ -measurable, and the fourth again from the law of iterated expectations. It follows that

$$\mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \mid \mathcal{G}_t \right] \cdot I_{A \cap B} \right] = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{1}{2} r' r \right) \cdot I_{A \cap B} \right].$$

□

Theorem F.3 (Girsanov's Theorem). *For any $r \in \mathbb{R}^3$,*

$$\mathbb{E}_t^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \mid s_{t+1} \right] = \exp \left(\frac{1}{2} r' r \right)$$

with probability equal to 1.

Proof. $\mathbb{E}_t^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \mid s_{t+1} \right]$ is the $\mathbb{Q}$ -conditional expectation of $\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right)$ with respect to $\mathcal{G}_t$, so it is trivially measurable with respect to $\mathcal{G}_t$. Likewise, $\exp \left(\frac{1}{2} r' r \right)$, being a non-random scalar, is trivially $\mathcal{G}_t$ -measurable.

Define

$$J = \{A \cap B \mid A \in \mathcal{F}_t, B \in \sigma(s_{t+1})\},$$

where σs_{t+1} is the σ -algebra generated by the random variable s_{t+1} . A well-known result in measure theory tells us that J is a π -system, that is, a collection of sets that is closed under intersections, that generates the σ -algebra $\mathcal{G}_t$ ⁴.

Define the functions $\pi : \mathcal{G}_t \rightarrow [0, +\infty]$ and $v : \mathcal{G}_t \rightarrow [0, +\infty]$ as

$$\begin{aligned} \pi(H) &= \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \mid s_{t+1} \right] \cdot I_H \right] \\ &:= \int_H \mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \mid s_{t+1} \right] d\mathbb{Q} \\ v(H) &= \mathbb{E}^{\mathbb{Q}} \left[\exp\left(\frac{1}{2} r' r\right) \cdot I_H \right] \\ &:= \exp\left(\frac{1}{2} r' r\right) \cdot \int_H 1 d\mathbb{Q} \end{aligned}$$

for any $H \in \mathcal{G}_t$. These functions are well-defined for each $H \in \mathcal{G}_t$ because the conditional expectation $\mathbb{E}_t^{\mathbb{Q}} \left[\exp(-r' v_{t+1}^{\mathbb{Q}}) \mid s_{t+1} \right]$ is a $\mathcal{G}_t$ -measurable real valued function and $\exp\left(\frac{1}{2} r' r\right)$ is a non-random constant. Heuristically, π and v are the indefinite integrals

⁴The proof of this result is straightforward. First, it is trivial to show that J is a π -system.

To show that the σ -algebra generated by J is exactly the union σ -algebra $\mathcal{G}_t$, we first note that $J \subset \mathcal{G}_t$. For any $A \in \mathcal{F}_t$ and $B \in \sigma(s_{t+1})$,

$$A \in \mathcal{F}_t \cup \sigma(s_{t+1}) \subset \mathcal{G}_t$$

and likewise for B . Since σ -algebras like $\mathcal{G}_t$ are closed under finite intersections, $A \cap B \in \mathcal{G}_t$. $\mathcal{G}_t$ is a σ -algebra containing J and the σ -algebra σJ generated by J is the smallest σ -algebra containing J , so it follows that $\sigma J \subset \mathcal{G}_t$.

For the reverse inclusion, note that, for any $A \in \mathcal{F}_t$, because the sample space Ω is an element of $\sigma(s_{t+1})$ by the definition of σ -algebras, we have

$$A = A \cap \Omega \in J.$$

In other words, $\mathcal{F}_t \subset J$, and by a similar line of reasoning, $\sigma(s_{t+1}) \subset J$ as well. Therefore, $\mathcal{F}_t \cup \sigma(s_{t+1}) \subset J \subset \sigma J$, and it follows that $\mathcal{G}_t$, the smallest σ -algebra containing $\mathcal{F}_t \cup \sigma(s_{t+1})$, is also contained in σJ . This establishes the result.

of $\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}]$ and $\exp(\frac{1}{2}r'r)$ with respect to the measure $\mathbb{Q}$, and they are well-defined measures on the measurable space $(\Omega, \mathcal{G}_t)$ by Proposition I.5.6 in Cinlar (2011).

It is easy to see that $v(\Omega) = \exp(\frac{1}{2}r'r)$. As for the measure π , note first that

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}]] = \mathbb{E}^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}})] = \mathbb{E}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}})]]$$

by the law of iterated expectations. By implication,

$$\begin{aligned} \pi(\Omega) &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}]] \\ &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}})]] \\ &= \exp\left(\frac{1}{2}r'r\right), \end{aligned}$$

where the last equality follows from putting $B = \Omega$ in Lemma D.1. This shows us that, π and v are finite measures on $(\Omega, \mathcal{G}_t)$ with the same total measure.

We have shown so far that J is a π -system that generates $\mathcal{G}_t$, and that π and v are finite measures on $(\Omega, \mathcal{G}_t)$ such that $\pi(\Omega) = v(\Omega)$. Furthermore, Lemma D.2 shows us that $\pi(H) = v(H)$ for any $H \in J$. Therefore, by Proposition I.3.7 in Cinlar (2011), $\pi = v$ on $\mathcal{G}_t$, that is,

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\exp(-r'v_{t+1}^{\mathbb{Q}}) \mid s_{t+1}] \cdot I_H] = \mathbb{E}^{\mathbb{Q}}[\exp(\frac{1}{2}r'r) \cdot I_H]$$

for any $H \in \mathcal{G}_t$. Now the definition of conditional expectations (see, for example, Definition IV.1.4 in Cinlar (2011)) and the uniqueness of conditional expectations up to

almost sure equality imply that, with probability equal to 1,

$$\mathbb{E}_t^{\mathbb{Q}} \left[\exp \left(-r' v_{t+1}^{\mathbb{Q}} \right) \mid s_{t+1} \right] = \exp \left(\frac{1}{2} r' r \right).$$

□

G Essentially Affine Market Prices of Risk and Risk Premia

We first derive the market prices of risk; the exposition follows Cheridito, Filipovic, and Kimmel (2007) almost exactly.

Our assumed risk-neutral and physical dynamics imply that

$$\begin{aligned} \mathbb{E}_t^{\mathbb{Q}} [f_{t+1} \mid s_{t+1}] &= K_{s_{t+1}}^{\mathbb{Q}} + G^{\mathbb{Q}} f_t \\ \mathbb{E}_t [f_{t+1} \mid s_{t+1}] &= K_{s_{t+1}}^{\mathbb{Q}} + G^{\mathbb{Q}} f_t + \Omega_{s_{t+1}}^{\frac{1}{2}} \cdot \lambda_{t,s_{t+1}}. \end{aligned}$$

Therefore,

$$\Omega_{s_{t+1}}^{\frac{1}{2}} \cdot \lambda_{t,s_{t+1}} = \mathbb{E}_t [f_{t+1} \mid s_{t+1}] - \mathbb{E}_t^{\mathbb{Q}} [f_{t+1} \mid s_{t+1}],$$

and $\lambda_{t,s_{t+1}}$ is given by

$$\lambda_{t,s_{t+1}} = \left(\Omega_{s_{t+1}}^{\frac{1}{2}} \right)^{-1} \left(\mathbb{E}_t [f_{t+1} \mid s_{t+1}] - \mathbb{E}_t^{\mathbb{Q}} [f_{t+1} \mid s_{t+1}] \right),$$

as a function of the factor risk premium $\mathbb{E}_t [f_{t+1} \mid s_{t+1}] - \mathbb{E}_t^{\mathbb{Q}} [f_{t+1} \mid s_{t+1}]$. Now make use of the physical dynamics of the factors:

$$\mathbb{E}_t [f_{t+1} \mid s_{t+1}] = \mathbb{E}_t \left[G_{s_{t+1}}^{\mathbb{P}} f_t + \Omega_{s_{t+1}}^{\frac{1}{2}} \cdot v_{t+1}^{\mathbb{P}} \right] = G_{s_{t+1}}^{\mathbb{P}} f_t.$$

As such, the market prices of risk can be rewritten in terms of the factors f_t as

$$\begin{aligned}\lambda_{t,s_{t+1}} &= \left(\Omega_{s_{t+1}}^{\frac{1}{2}}\right)^{-1} \left(G_{s_{t+1}}^{\mathbb{P}} f_t - K_{s_{t+1}}^{\mathbb{Q}} - G^{\mathbb{Q}} f_t\right) \\ &= \left(\Omega_{s_{t+1}}^{\frac{1}{2}}\right)^{-1} \left[-K_{s_{t+1}}^{\mathbb{Q}} + (G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}}) f_t\right].\end{aligned}$$

We have thus seen that the market price of risk is determined as follows:

$$\begin{aligned}\lambda_{t,s_{t+1}} &= \lambda_{s_{t+1}} + \Lambda_{s_{t+1}} \cdot f_t, \\ \lambda_{s_{t+1}} &= -\left(\Omega_{s_{t+1}}^{\frac{1}{2}}\right)^{-1} K_{s_{t+1}}^{\mathbb{Q}} \\ \Lambda_{s_{t+1}} &= \left(\Omega_{s_{t+1}}^{\frac{1}{2}}\right)^{-1} \left(G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}}\right).\end{aligned}$$

Next, using the discrete version of Girsanov's theorem and the essentially affine form of $\lambda_{t,s_{t+1}}$ above, we derive an expression for the one-period ahead expected excess return on zero-coupon bonds.

Recall that the one-period ahead holding period return for a τ -maturity bond at time t is given as

$$hpr_{t+1}^{(\tau)} = \log(P_{t+1}(\tau - 1)) - \log(P_t(\tau)).$$

Since bond prices are given as exponential-affine functions of f_t and f_{t+1} has a VAR(1) representation under the risk-neutral measure, the holding period return can be written in terms of time t factors as

$$\begin{aligned}hpr_{t+1}^{(\tau)} &= -a_{s_{t+1}}(\tau - 1) - b(\tau - 1)' f_{t+1} + a_{s_t}(\tau) + b(\tau)' f_t \\ &= -a_{s_{t+1}}(\tau - 1) - b(\tau - 1)' K_{s_{t+1}}^{\mathbb{Q}} - b(\tau - 1)' G^{\mathbb{Q}} f_t - b(\tau - 1)' \Omega_{s_{t+1}}^{\frac{1}{2}} v_{t+1}^{\mathbb{Q}} \\ &\quad + a_{s_t}(\tau) + b(\tau)' f_t.\end{aligned}$$

The intercept and factor loadings $a_{s_t}(\tau)$ and $b(\tau)$ were seen to satisfy the recursions

$$\begin{aligned} a_{s_t}(\tau) &= \mathbb{E}_t^{\mathbb{Q}} \left[a_{s_{t+1}}(\tau-1) + b(\tau-1)' K_{s_{t+1}}^{\mathbb{Q}} - \frac{1}{2} b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right] \\ b(\tau) &= \beta + G^{\mathbb{Q}} b(\tau-1) \end{aligned}$$

if we ignore Jensen's inequality and interchange logs and $\mathbb{Q}$ -expectations. Substituting these equations into the equation for holding period returns, we have

$$\begin{aligned} hpr_{t+1}^{(\tau)} &= \beta' f_t + \mathbb{E}_t \left[a_{s_{t+1}}(\tau-1) - b(\tau-1)' K_{s_{t+1}}^{\mathbb{Q}} - \frac{1}{2} b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right] \\ &\quad - a_{s_{t+1}}(\tau-1) - b(\tau-1)' K_{s_{t+1}}^{\mathbb{Q}} - b(\tau-1)' \Omega_{s_{t+1}}^{\frac{1}{2}} v_{t+1}^{\mathbb{Q}}, \end{aligned}$$

where the $\mathbb{Q}$ -expectations have been replaced with $\mathbb{P}$ -expectations because the distribution of s_{t+1} given s_t is the same across the risk-neutral and physical measures. As such, we have

$$\begin{aligned} \mathbb{E}_t \left[hpr_{t+1}^{(\tau)} \mid s_{t+1} \right] &= r_t - \mathbb{E}_t \left[\frac{1}{2} b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right] - b(\tau-1)' \Omega_{s_{t+1}}^{\frac{1}{2}} \lambda_{t,s_{t+1}} \\ &= r_t - \mathbb{E}_t \left[\frac{1}{2} b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right] \\ &\quad + b(\tau-1)' K_{s_{t+1}}^{\mathbb{Q}} - b(\tau-1)' \left(G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}} \right) f_t, \end{aligned}$$

where we used the fact that the mean of $v_{t+1}^{\mathbb{Q}}$ is equal to the market price of risk $\lambda_{t,s_{t+1}}$ given the information up to time t and s_{t+1} , as per Girsanov's theorem.

Finally, taking time t $\mathbb{P}$ -expectations on both sides yields, by the law of iterated expectations,

$$\begin{aligned} exr_{t+1}^{(\tau)} &= \mathbb{E}_t \left[hpr_{t+1}^{(\tau)} \right] - r_t \\ &= -\frac{1}{2} \mathbb{E}_t \left[b(\tau-1)' \Omega_{s_{t+1}} b(\tau-1) \right] + b(\tau-1)' \mathbb{E}_t \left[K_{s_{t+1}}^{\mathbb{Q}} \right] \end{aligned}$$

$$-b(\tau - 1)' \mathbb{E}_t \left[\left(G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}} \right) f_t \right].$$

H Details on Estimation

Here we present further details about our estimation method.

H.1 Rationale for Bayesian Estimation

We choose to sample the model parameters via Bayesian Gibbs sampling. This is because imposing the appropriate prior distributions on $G_{s_t}^{\mathbb{P}}$, among other parameters, can help stabilize estimates of the term premium. We can see from the formula for the term premium that the difference $G_{s_t}^{\mathbb{P}} - G^{\mathbb{Q}}$ between the mean reversion parameters of the physical and risk-neutral factor dynamics determines how the factors load on the term premium. Choosing the appropriate location and tightness of the prior distribution of $G_{s_t}^{\mathbb{P}}$ can thus prevent the estimated term premium from behaving too erratically. Indeed, by imposing a relatively tight prior on $g_1^{\mathbb{P}}$, we can smooth the sudden spike in the term premium around 2008 found in the term premium estimate of Christensen and Rudebusch (2016).

H.2 Data

Our sample includes the short rate and U.S. yields for $m = 17$ maturities: 3 and 6 months, and yearly maturities from 1 to 15 years. These maturities are chosen to coincide with those in Bauer and Rudebusch (2020). The yields with maturities one year or higher are selected as end-of-month yields in the Gurkaynak, Sack, and Wright (2007) (GSW) database. The 3-month and 6-month yields are taken from the H-15 Federal Release because the GSW database is unreliable at the short end of the yield curve. Following Chib and Ergashev (2009), we use the effective federal funds rate as a proxy for the 1-

month yield, which serves as our short rate r_t . To avoid issues with potential structural changes during the Great Inflation period from 1965 to 1982, we limit our sample to the period from January 1985 to April 2023.

Letting $\tau_1, \dots, \tau_m$ be the maturities in our sample apart from the short rate, the sample of time t yields of maturity 3 months or higher are collected in the vector

$$\mathcal{Y}_t = \begin{pmatrix} Y_t(\tau_1) \\ \vdots \\ Y_t(\tau_m) \end{pmatrix}.$$

The model-implied yield formula is

$$\mathcal{Y}_t = \mathcal{A}_{s_t} + \mathcal{B}f_t,$$

where

$$\mathcal{A}_{s_t} = \begin{pmatrix} \frac{a_{s_t}(\tau_1)}{\tau_1} \\ \vdots \\ \frac{a_{s_t}(\tau_m)}{\tau_m} \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} \frac{b(\tau_1)'}{\tau_1} \\ \vdots \\ \frac{b(\tau_m)'}{\tau_m} \end{pmatrix} \in \mathbb{R}^{m \times 3}.$$

H.3 Parameter and Factor Identification

Before estimating any term structure model, we must ensure that it is identified. In the term structure literature, it is customary to impose identification restrictions that insulates a model from invariant affine transformations (e.g. Dai and Singleton (2000), Joslin, Singleton, and Zhu (2011), and Hamilton and Wu (2012)). We impose the following identification restrictions:

- **Restrictions on Short Rate Dynamics** We require the intercept term in the

short rate equation to be 0 in both regimes, and that the level factor load with unity on the short rate.

- **Restrictions on $G^{\mathbb{Q}}$** We require the mean reversion parameter $G^{\mathbb{Q}}$ under the risk-neutral measure to be of the lower triangular form (B.2) for some $\kappa^{\mathbb{Q}} \in (0, +\infty)$.
- **Restrictions on Ω_{s_t}** We require that $\omega_{s_t}^2$ and Ω_{22} are the only non-zero entries in the factor error covariance matrix Ω_{s_t} , and that Ω_{22} is positive definite.
- **Restrictions on the Physical Factor Dynamics** We require that the slope and curvature factors $(S_t, C_t)'$ have unconditional mean zero under the physical measure.

The last restriction is taken from Hamilton and Wu (2012), where it is stated that the intercept of the factors under either the risk-neutral or the physical measure must be put equal to 0. Below we show that there are no invariant transformations of the factors f_t that satisfy the above restrictions simultaneously. Hamilton and Wu (2012) point out that, if both the physical and risk-neutral dynamics of the factors contain a unit root, then the model, specifically the intercept of the factor dynamics, is unidentified. In our model, while the physical factor dynamics are I(1) during the normal regime by design, the model is nonetheless identified because of our restrictions on the short rate dynamics.

Suppose that f_t is a factor process that satisfies the identification restrictions, so that

$$f_t = K_{s_t}^{\mathbb{Q}} + \underbrace{\begin{pmatrix} 1 & 1-x & -x \log(x) \\ 0 & x & x \log(x) \\ 0 & 0 & x \end{pmatrix}}_{G^{\mathbb{Q}}(x)} f_{t-1} + \underbrace{\begin{pmatrix} \omega_{s_t} & O_{1 \times 2} \\ O_{2 \times 1} & \Omega_{22}^{\frac{1}{2}} \end{pmatrix}}_{\Omega_{s_t}^{\frac{1}{2}}} v_t^{\mathbb{Q}}$$

for some $x \in (0, 1)$,

$$f_t = \begin{pmatrix} \rho_{s_t} & g_{s_t}^{\mathbb{P}'} \\ O_{2 \times 1} & \tilde{G} \end{pmatrix} f_{t-1} + \Omega_{s_t}^{\frac{1}{2}} v_t^{\mathbb{P}},$$

and

$$r_t = \beta' f_t.$$

Let

$$\hat{f}_t = A + B f_t$$

be an invariant affine transformation of f_t that also satisfies the above restrictions, that is, there exists a $y \in (0, 1)$ such that

$$\hat{f}_t = \underbrace{\hat{K}_{s_t}^{\mathbb{Q}} + \begin{pmatrix} 1 & 1-y & -y \log(y) \\ 0 & y & y \log(x) \\ 0 & 0 & y \end{pmatrix}}_{G^{\mathbb{Q}}(y)} \hat{f}_{t-1} + \underbrace{\begin{pmatrix} \sigma_{s_t} & O_{1 \times 2} \\ O_{2 \times 1} & \Sigma_{22}^{\frac{1}{2}} \end{pmatrix}}_{\Sigma_{s_t}^{\frac{1}{2}}} v_t^{\mathbb{Q}},$$

$r_t = \beta' \hat{f}_t$, and the unconditional means of the last two elements of $\hat{f}_t$ are 0.

We first show that $B = a \cdot I_3$ for some $a \neq 0$. Using the relationship

$$f_t = B^{-1} (\hat{f}_t - A),$$

the $\mathbb{Q}$ -dynamics of $\hat{f}_t$ are given by

$$\hat{f}_t = A + B f_t$$

$$\begin{aligned}
&= A + BK_{s_t}^{\mathbb{Q}} + BG^{\mathbb{Q}}(x)f_{t-1} + B\Omega_{s_t}^{\frac{1}{2}}v_t^{\mathbb{Q}} \\
&= BK_{s_t}^{\mathbb{Q}} + \left(I_3 - BG^{\mathbb{Q}}(x)B^{-1}\right)A \\
&\quad + BG^{\mathbb{Q}}(x)B^{-1} \cdot \hat{f}_{t-1} + B\Omega_{s_t}^{\frac{1}{2}}v_t^{\mathbb{Q}}.
\end{aligned}$$

It follows that

$$BG^{\mathbb{Q}}(x) = G^{\mathbb{Q}}(y)B$$

and

$$B\Omega_{s_t}B' = \Sigma_{s_t}.$$

Let B be given as

$$B = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

The equality

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} 1 & 1-x & -x\log(x) \\ 0 & x & x\log(x) \\ 0 & 0 & x \end{pmatrix} = \begin{pmatrix} 1 & 1-y & -y\log(y) \\ 0 & y & y\log(x) \\ 0 & 0 & y \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

yields the 9 equations

$$\begin{pmatrix} d(1-y) = gy\log(y) & (b-a)(1-x) + e(1-y) = hy\log(y) & e(1-x) + f(1-y) - iy\log(y) = (b-a)x\log(x) \\ d(1-y) = gy\log(y) & e(x-y) + d(1-x) = hy\log(y) & f(y-x) + iy\log(y) = (e-d)x\log(x) \\ gy = g & g(1-x) = h(y-x) & i(y-x) = (h-y)x\log(x) \end{pmatrix}.$$

The first column of equations tell us that $g = 0$, since $y < 1$, which implies that $d = 0$.

Next, from the second column we can tell that $h = 0$ and $x = y$, and as such that $e = a - b$, where $a \neq b$ and $a \neq 0$ because B is nonsingular. Finally, the last column tells us that

$e = i$ and $f = -c$. Together, B can be simplified as

$$B = \begin{pmatrix} a & b & c \\ 0 & a-b & -c \\ 0 & 0 & a-b \end{pmatrix}.$$

To pin down the values of b and c , we now turn to the equation $B\Omega_{s_t}B' = \Sigma_{s_t}$. Substituting the expression for B derived above into the equation, we obtain the following equations:

$$\begin{aligned} a^2\omega_{s_t}^2 &= \sigma_{s_t}^2, \\ \begin{pmatrix} b & c \end{pmatrix} \Omega_{22} \begin{pmatrix} a-b & 0 \\ -c & a-b \end{pmatrix} &= O_{1 \times 2}, \\ \begin{pmatrix} a-b & -c \\ 0 & a-b \end{pmatrix} \Omega_{22} \begin{pmatrix} a-b & 0 \\ -c & a-b \end{pmatrix} &= \Sigma_{22}. \end{aligned}$$

Because both Ω_{22} and $\begin{pmatrix} a-b & 0 \\ -c & a-b \end{pmatrix}$ are nonsingular, the second equation tells us that $b = c = 0$, while the first implies that $a \neq 0$. Therefore, $B = a \cdot I_3$ for some $a \neq 0$.

Next, we show that A must be a zero vector and that the scale factor a equals 1. To this end, note that the short rate dynamics under $\hat{f}_t$ equals

$$r_t = \beta' f_t = -\beta' B^{-1} A + \beta' B^{-1} \cdot \hat{f}_t.$$

This tells us that $\beta' A = 0$ for both $s_t = 1, 2$. In addition, the intercept of the factor dynamics of $\hat{f}_t$ under the physical measure must equal

$$(I_3 - BG_{s_t}^{\mathbb{P}} B^{-1}) A = (I_3 - G_{s_t}^{\mathbb{P}}) A.$$

In particular, the last two elements are given as

$$(I_2 - \tilde{G}) A^{(2)},$$

where $A^{(2)}$ collects the last two elements of A . By the identification restriction on the unconditional mean of the slope and curvature factors, the above 2×1 vector must equal 0 for any time t , and since the eigenvalues of $\tilde{G}$ are all contained within the unit circle, $A^{(2)} = O_{2 \times 1}$. Finally, since the first element of both β and $\beta' B^{-1}$ must equal 1, $a = 1$ and the condition $\beta' A = 0$ requires $A = O_{3 \times 1}$.

We have thus established that $A = O_{3 \times 1}$ and $B = I_3$, meaning that $\hat{f}_t = f_t$. Therefore, f_t collects the unique factors that satisfy our identification restrictions, and the model is identified.

H.4 The State-Space Representation of the Model

To complete the model, we assume the presence of a short rate measurement error u_t with a variance of σ_u^2 , and a vector of measurement errors e_t pertaining to the other sample yields. As is standard in the literature, we assume the measurement errors for the yields are mutually independent and have variances $\sigma_{e,1}^2, \dots, \sigma_{e,m}^2$. Our model is thus comprised of the measurement equations

$$r_t = \beta' f_t + \sigma_u \cdot u_t$$

$$\mathcal{Y}_t = \mathcal{A}_{s_t} + \mathcal{B} f_t + \Sigma_e^{\frac{1}{2}} \cdot e_t,$$

where $\Sigma_e = \text{diag}(\sigma_{e,1}^2, \dots, \sigma_{e,m}^2)$, and the transition equation

$$f_t = \underbrace{\begin{pmatrix} \rho_{s_t} & g_{s_t}^{\mathbb{P}'} \\ O_{2 \times 1} & \tilde{G} \end{pmatrix}}_{G_{s_t}^{\mathbb{P}}} f_{t-1} + \Omega_{s_t}^{\frac{1}{2}} \cdot v_t^{\mathbb{P}}.$$

In the estimation process, all yields are appropriately scaled via the scaling parameter $c = 1200$ to transform them into per annum percentage values.

We follow Hamilton and Wu (2012), as well as the RY specification of case **P** in Joslin, Singleton, and Zhu (2011), to assume that the short rate, and the 3 year and 10 year yields are observed without error. These three yields are now related to the factors as

$$\begin{pmatrix} r_t \\ Y_t(36) \\ Y_t(120) \end{pmatrix} = \begin{pmatrix} 0 \\ a_{s_t}(36)/36 \\ a_{s_t}(120)/120 \end{pmatrix} + \begin{pmatrix} \beta' \\ b(36)'/36 \\ b(120)'/120 \end{pmatrix} f_t.$$

Here, since the federal funds rate, being an overnight rate, is close to a zero maturity yield, we approximate the factor loadings β as $(1, 0, 0)'$, essentially identifying the level factor with the short rate. The first element in $b(\tau)'/\tau$ equals 1 for any maturity τ , so the equations for the 3- and 10-year yields can be reformulated in terms of their yield spreads as

$$\underbrace{\begin{pmatrix} Y_t(36) - r_t \\ Y_t(120) - r_t \end{pmatrix}}_{\mathcal{Y}S_t^{(1)}} = \underbrace{\begin{pmatrix} a_{s_t}(36)/36 \\ a_{s_t}(120)/120 \end{pmatrix}}_{\mathcal{A}_{s_t}^{(1)}} + \underbrace{\begin{pmatrix} b(36)'/36 \\ b(120)'/120 \end{pmatrix}}_{\mathcal{B}^{(1)}} f_t^{SC},$$

where f_t^{SC} collects the slope and curvature factors S_t, C_t . The rest of the yield spreads

are collected as

$$\underbrace{\begin{pmatrix} Y_t(\tau_3) - r_t \\ \vdots \\ Y_t(\tau_m) - r_t \end{pmatrix}}_{\mathcal{Y}\mathcal{S}_t^{(2)}} = \underbrace{\begin{pmatrix} a_{s_t}(\tau_3)/\tau_3 \\ \vdots \\ a_{s_t}(\tau_m)/\tau_m \end{pmatrix}}_{\mathcal{A}_{s_t}^{(2)}} + \underbrace{\begin{pmatrix} b(\tau_3)'/\tau_3 \\ \vdots \\ b(\tau_m)'/\tau_m \end{pmatrix}}_{\mathcal{B}^{(2)}} f_t^{SC} + \Sigma_e^{\frac{1}{2}} \cdot e_t,$$

where e_t is now an $(m-2)$ -dimensional random vector of standard normally distributed measurement errors, and Σ_e likewise collects the measurement errors of the $m-2$ yields that are observed with error.

The modified state-space representation of the model is now given in terms of the yield spreads as

$$\begin{pmatrix} \mathcal{Y}\mathcal{S}_t^{(1)} \\ \mathcal{Y}\mathcal{S}_t^{(2)} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{s_t}^{(1)} \\ \mathcal{A}_{s_t}^{(2)} \end{pmatrix} + \begin{pmatrix} \mathcal{B}^{(1)} \\ \mathcal{B}^{(2)} \end{pmatrix} f_t^{SC} + \begin{pmatrix} O_{2 \times (m-2)} \\ \Sigma_e^{\frac{1}{2}} \end{pmatrix} e_t \quad (\text{H.1})$$

$$f_t = G_{s_t}^{\mathbb{P}} f_{t-1} + \Omega_{s_t}^{\frac{1}{2}} \cdot v_t^{\mathbb{P}} \quad (\text{H.2})$$

We also assume that the regime in the first period is 1, that is, $s_1 = 1$, and that the Markov process $\{s_t\}_{t \in \mathbb{Z}}$ is irreducible and ergodic, so that it has a stationary distribution η^5 .

H.5 Estimation Strategy

Hamilton and Wu (2012) point out that the state-space model in Equations (H.1) and (H.2) is observationally equivalent to a restricted Gaussian VAR. In our case, the re-

⁵We assume that the process has arrived at the stationary distribution η by the time the sample starts. Thus, the unconditional distribution of s_0 is η under both the physical and risk-neutral measures.

stricted Gaussian VAR system is characterized by the following three equations:

$$r_t = -g_{st}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \mathcal{A}_{st-1}^{(1)} + \rho_{st} r_{t-1} + g_{st}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \cdot \mathcal{Y} \mathcal{S}_{t-1}^{(1)} + \omega_{st} \cdot v_{1t}^{\mathbb{P}} \quad (\text{H.3})$$

$$\mathcal{Y} \mathcal{S}_t^{(1)} = \left(\mathcal{A}_{st}^{(1)} - \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{A}_{st-1}^{(1)} \right) + \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{Y} \mathcal{S}_{t-1}^{(1)} + \mathcal{B}^{(1)} \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}} \quad (\text{H.4})$$

$$\mathcal{Y} \mathcal{S}_t^{(2)} = \left(\mathcal{A}_{st}^{(2)} - \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \mathcal{A}_{st}^{(1)} \right) + \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \cdot \mathcal{Y} \mathcal{S}_t^{(1)} + \Sigma_e^{\frac{1}{2}} \cdot e_t, \quad (\text{H.5})$$

where $v_{SC,t}^{\mathbb{P}}$ collects the last two elements of $v_t^{\mathbb{P}}$. The first two equations represent the dynamic part of the system and the third the static part. The three equations above can be taken together to form the larger VAR system

$$X_t = C_{1,st,s_{t-1}} \cdot \mathcal{A} + \Phi_{st} \cdot X_{t-1} + \Sigma_{st} \cdot u_t, \quad (\text{H.6})$$

where

$$X_t = \left(r_t, \mathcal{Y} \mathcal{S}_t^{(1)}, \mathcal{Y} \mathcal{S}_t^{(2)} \right) \quad (\text{H.7})$$

and $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$. The parameters $C_{1,st,s_{t-1}}$, Φ_{st} and Σ_{st} are defined in Appendix I.

H.6 Relationship between Structural and Reduced Form Parameters

There are a total of 32 structural parameters to be estimated, collected as

$$\theta = \{\rho_2, g_1^{\mathbb{P}}, \omega_1^2, \omega_2^2, \tilde{G}, \Omega_{22}, \kappa^{\mathbb{Q}}, K_1^{\mathbb{Q}}, K_2^{\mathbb{Q}}, \Sigma_e, P\}$$

alongside the regime process $S = \{s_t\}_{1 \leq t \leq T}$. Estimating the $\mathbb{Q}$ -parameter $K_{st}^{\mathbb{Q}}$ has traditionally proven difficult. In particular, we wish to draw attention to how the intercept

term $\mathcal{A}_{s_t} = (\mathcal{A}_{s_t}^{(1)}, \mathcal{A}_{s_t}^{(2)})$ is defined as

$$\mathcal{A}_{s_t} = -\frac{1}{2}c_{0,s_t} + c_{1,s_t} \begin{pmatrix} K_1^{\mathbb{Q}} \\ K_2^{\mathbb{Q}} \end{pmatrix}, \quad (\text{H.8})$$

where

$$c_{0,s_t} = \begin{pmatrix} \frac{1}{\tau_1} \sum_{i=1}^{\tau_1-1} b(\tau_1 - i)' \left(\sum_{j=1}^N \Omega_j \cdot P_{s_t,j}^{(i)} \right) b(\tau_1 - i) \\ \vdots \\ \frac{1}{\tau_m} \sum_{i=1}^{\tau_m-1} b(\tau_m - i)' \left(\sum_{j=1}^N \Omega_j \cdot P_{s_t,j}^{(i)} \right) b(\tau_m - i) \end{pmatrix} \quad (\text{H.9})$$

$$c_{1,s_t} = \begin{pmatrix} \frac{1}{\tau_1} \sum_{i=1}^{\tau_1-1} b(\tau_1 - i) \cdot P_{s_t,1}^{(i)} & \frac{1}{\tau_1} \sum_{i=1}^{\tau_1-1} b(\tau_1 - i) \cdot P_{s_t,2}^{(i)} \\ \vdots & \vdots \\ \frac{1}{\tau_m} \sum_{i=1}^{\tau_m-1} b(\tau_m - i) \cdot P_{s_t,1}^{(i)} & \frac{1}{\tau_m} \sum_{i=1}^{\tau_m-1} b(\tau_m - i) \cdot P_{s_t,2}^{(i)} \end{pmatrix}. \quad (\text{H.10})$$

While $\mathcal{A}_{s_t}$, and by extension the measurement equation, are affine in $K_1^{\mathbb{Q}}$ and $K_2^{\mathbb{Q}}$, Ω_{s_t} and P enter into the measurement equation in a non-linear manner through c_{0,s_t} . This renders estimation difficult, with maximum likelihood methods requiring non-linear numerical optimization and Bayesian methods requiring the use of the Metropolis-Hasting (MH) algorithm.

Instead of directly sampling from the structural parameters, a simpler way is to take a two-step approach, as in Hamilton and Wu (2012) and many other papers that employ minimum chi-square estimation. First, we sample the reduced form parameters

$$\theta^* = \{\rho_2, g_1^{\mathbb{P}}, \omega_1^2, \omega_2^2, \tilde{G}, \Omega_{22}, \kappa^{\mathbb{Q}}, \mathcal{A}_1, \mathcal{A}_2, \Sigma_e, P\}$$

that are present in the system in equation (H.7). Afterward, we can recover the structural

parameter $K_{st}^{\mathbb{Q}}$ by applying the transformation

$$\begin{pmatrix} K_1^{\mathbb{Q}} \\ K_2^{\mathbb{Q}} \end{pmatrix} = [c'_{11}c_{11} + c'_{12}c_{12}]^{-1} \left[c'_{11} \left(\mathcal{A}_1 + \frac{1}{2}c_{01} \right) + c'_{12} \left(\mathcal{A}_2 + \frac{1}{2}c_{02} \right) \right]. \quad (\text{H.11})$$

As shown in Hamilton and Wu (2012), among others, this two-step procedure generates consistent and asymptotically normal estimators of the structural parameters. In light of the class of Bernstein von-Mises (BvM) theorems, we expect the posterior distribution of the structural parameters obtained in this manner to possess similar asymptotic properties.

When it comes to Bayesian estimation, estimating θ^* instead of θ confers an additional advantage. In short, Gibbs sampling $\mathcal{A}_{st}$ instead of $K_{st}^{\mathbb{Q}}$ can help facilitate the choice of prior distributions. With $\mathcal{A}_{st}$, since the unconditional mean of the slope and curvature parameters is 0, we know that $\mathcal{A}_{st}$ collects the regime-wise unconditional mean of the yield spreads $\mathcal{YS}_t^{(1)}$ and $\mathcal{YS}_t^{(2)}$. In contrast, all we know about $K_{st}^{\mathbb{Q}}$ is that it helps determine $\mathcal{A}_{st}$, but the specific role it plays in determining the level of the yield spreads is obfuscated by the presence of the Jensen's inequality term, among others, in the formula for $\mathcal{A}_{st}$. Therefore, by choosing to sample $\mathcal{A}_{st}$, we can impose reasonable prior distributions centered around the sample average of the yield spreads.

H.7 Sampling Algorithm

The following is the rough outline of our block-wise Gibbs sampling algorithm. It essentially samples each parameter by successively cycling through multiple linear regression models, so that we need only rely on the simple Normal-Normal and Inverse Gamma-Inverse Gamma (or Inverse Wishart-Inverse Wishart) updates. Details are provided in Appendix I.

Block 1 : Sampling $\rho_2, g_1^{\mathbb{P}}, \omega_1^2, \omega_2^2$

These parameters only appear in Equation (H.3), which can be written as

$$r_t = \rho_{s_t} r_{t-1} + g_{s_t}^{\mathbb{P}'} f_{t-1}^{SC} + \omega_{s_t} \cdot v_{1t}^{\mathbb{P}}$$

in terms of the factors, which are known given $\kappa^{\mathbb{Q}}$ and $\mathcal{A}$.

Therefore, we can sample the innovation variances ω_1^2, ω_2^2 via the Inverse Gamma-Inverse Gamma update, while $\bar{r}_2$ and $g_1^{\mathbb{P}}$ can be sampled via the Normal-Normal update.

Block 2: Sampling $\tilde{G}, \Omega_{22}$

These parameters only appear in Equation (H.4), which can be written in terms of the factors as

$$f_t^{SC} = \tilde{G} f_{t-1}^{SC} + \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}}.$$

Therefore, Ω_{22} and $\tilde{G}$ can be sampled via the Inverse Wishart-Inverse Wishart and Normal-Normal updates, as the innovation variance and coefficient from a standard Seemingly Unrelated Regression (SUR) model.

Block 3: Sampling Σ_e

The measurement error variances only appear in Equation (H.5):

$$\mathcal{Y}S_t^{(2)} = \left(\mathcal{A}_{s_t}^{(2)} - \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \mathcal{A}_{s_t}^{(1)} \right) + \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \cdot \mathcal{Y}S_t^{(1)} + \Sigma_e^{\frac{1}{2}} \cdot e_t.$$

Each measurement error variance can therefore be sampled via the Inverse Gamma-Inverse Gamma update.

Block 4: Sampling P

The transition probability P appears nowhere among Equations (H.3) to (H.5).

Therefore, it can be Gibbs sampled via the usual Beta-Beta update.

Block 5: Sampling $\kappa^{\mathbb{Q}}$ and $\mathcal{A}$

These are parameters that appear in all three equations. $\mathcal{A}$ is part of the intercept term in Equation (H.6):

$$X_t = C_{1,s_t,s_{t-1}} \cdot \mathcal{A} + \Phi_{s_t} \cdot X_{t-1} + \Sigma_{s_t}^{\frac{1}{2}} \cdot u_t.$$

As such, it can be sampled via the Normal-Normal update process.

As for $\kappa^{\mathbb{Q}}$, we assume that it takes finitely many values, each corresponding to a integer-valued maturity peak ranging from 12 months to 60 months. This is similar to the grid search method employed to estimate $\kappa^{\mathbb{Q}}$ found in Diebold and Li (2006). Equation (H.6) is used to compute the posterior density of each of these values of $\kappa^{\mathbb{Q}}$, and then we sample $\kappa^{\mathbb{Q}}$ given this discrete distribution.

Block 6: Sampling S

Finally, we sample the regime process S by making use of the usual Hamilton filter and Carter-Kohn backward recursion.

H.8 Prior Distributions

We impose the following priors on the model parameters:

$$\omega_i^2 \sim \text{InverseGamma}\left(\frac{a_{i,0}}{2}, \frac{d_{i,0}}{2}\right) \quad \text{for } i = 1, 2,$$

$$g_1^{\mathbb{P}} \sim \mathcal{N}[\bar{g}_0, V_{g,0}],$$

$$\rho_2 \sim \mathcal{N}[\bar{\rho}_0, V_{\rho,0}],$$

$$\Omega_{22} \sim \text{InverseWishart}(v_0, \Psi_0),$$

$$\begin{aligned}
\text{vec}(\tilde{G}^{\mathbb{P}'}) &\sim \mathcal{N}[\bar{\gamma}_0, V_{\gamma,0}], \\
\mathcal{A} &\sim \mathcal{N}[\bar{\mathcal{A}}_0, V_{\mathcal{A},0}], \\
\sigma_{e,i}^2 &\sim \text{InverseGamma}\left(\frac{a_0^{(i)}}{2}, \frac{d_0^{(i)}}{2}\right) \quad \text{for } 1 \leq i \leq m-2, \\
P_{ii} &\sim \text{beta}(\alpha_{i,0}, \beta_{i,0}) \quad \text{for } i = 1, 2.
\end{aligned}$$

We assume the parameters are all mutually independent and independent of the regime process S , except for the transition probabilities P_{11} and P_{22} .

The prior distributions are designed to lead to conjugate posteriors. The hyperparameters are at times chosen based on the least squares estimates of the corresponding parameters, and at others to ensure efficient sampling, particularly for the hyperparameters that govern the priors of $G_{st}^{\mathbb{P}}$. In all cases, the prior means are chosen in a manner that emphasizes the unique role that each parameter plays in the model (e.g. $\mathcal{A}$ as the collection of the regime-wise yield spread means). The priors are mostly uninformative, except for that of the mean-reversion parameter ρ_2 , which, without a tight prior, tends to be sampled too close to 1.

H.9 Prior Hyperparameters

Unless mentioned otherwise, the prior variances are chosen so that the prior distribution is uninformative.

- **Prior Distribution of $g_1^{\mathbb{P}}$**

The prior mean of the coefficients on the slope and curvature factors in the equation describing the dynamics of the level factor during the normal regime is set as

$$\bar{g}_0 = \begin{pmatrix} 1 - \exp(-\hat{\kappa}) & \hat{\kappa} \exp(-\hat{\kappa}) \end{pmatrix}',$$

which collects the (1,2) and (1,3) entries of $G^{\mathbb{Q}}$.

- **Prior Distribution of $\tilde{G}^{\mathbb{P}}$**

The prior mean of $\tilde{G}^{\mathbb{P}}$, the lower 2×2 matrix in $G_{st}^{\mathbb{P}}$, is set as

$$\hat{G}_0 = \begin{pmatrix} \exp(-\bar{\kappa}_0) & -\hat{\kappa} \exp(-\hat{\kappa}) \\ 0 & \exp(-\hat{\kappa}), \end{pmatrix}$$

which is the lower 2×2 block of $G^{\mathbb{Q}}$. By implication, $\bar{\gamma}_0 = \text{vec}(\hat{G}_0')$.

- **Prior Distribution of ρ_2**

The prior mean of ρ_2 is chosen as the least squares estimate of ρ in the regression $r_t = \rho r_{t-1} + \omega_2 \cdot u_t$, computed using only the periods during which the short rate is smaller than 0.25. When there are no periods where the short rate is smaller than 0.25, we put the prior mean of ρ_2 equal to 0.9. To avoid ρ_2 from being sampled as a value that is too close to 1, we impose a tight prior standard deviation equal to 0.05. In particular, the prior mean of ρ_2 in the full sample is given as 0.78, so this ensures that 0.9 is just outside two standard deviations of the prior mean. Note that this is the only parameter on which we impose a tight prior.

- **Prior Distribution of Ω_{22}**

The slope and curvature factors S_t, C_t have a data generating process independent of the level factor, given as

$$\begin{pmatrix} S_t \\ C_t \end{pmatrix} = \tilde{G}^{\mathbb{P}} \begin{pmatrix} S_{t-1} \\ C_{t-1} \end{pmatrix} + \varepsilon_{2,t}^{\mathbb{P}}, \quad \varepsilon_{2,t}^{\mathbb{P}} \sim iid \mathcal{N}(O_{2 \times 1}, \Omega_{22}).$$

This is simply a VAR(1) process, and the factors S_t, C_t are assumed to be observable, so the hyper-parameters v_0, Ψ_0 are determined so that the prior mean of Ω_{22} equals the MLE of Ω_{22} with $\tilde{G}^{\mathbb{P}}$ put equal to $\bar{G}_0$.

- **Prior Distribution of ω_1^2**

ω_1^2 is the error variance of the stationary process $\Delta L_t - g_1^{\mathbb{P}'} f_{t-1}^{(2)}$ during the normal regime. Therefore, $a_{1,0}$ and $d_{1,0}$ are chosen so that the prior mean of ω_1^2 equals the variance of $L_t - L_{t-1} - g_1^{\mathbb{P}'} f_{t-1}^{(2)}$ given that $g_1^{\mathbb{P}}$ equals its prior mean $\bar{g}_0$.

- **Prior Distribution of ω_2^2**

ω_2^2 represents the conditional yield volatility of the short rate during the lower bound regime. In this context, ω_1^2 is the conditional yield volatility of the short rate during the normal regime. In Christensen (2013), it is shown that the conditional yield volatility of the 3-month yield differs by about 7 times between the normal and lower bound regimes. Based on this result, we choose $a_{2,0}$ and $d_{2,0}$ so that the prior mean of ω_1^2 is 7^2 times larger than that of ω_2^2 .

- **Prior Distribution of P**

It is known that the expected duration of regime i is given by $\frac{1}{1-\bar{P}_{ii}}$ for $i = 1, 2$ under the assumption that $\{s_t\}$ is a Markov process. Therefore, for $i = 1, 2$, the hyper-parameters $\alpha_{i,0}$ and $\beta_{i,0}$ are chosen so that the prior mean for P_{ii} , denoted $\bar{P}_{ii}$, satisfies

$$D_i = \frac{1}{1 - \bar{P}_{ii}},$$

where D_i is the average duration of regime i under the heuristic regime process $\bar{S}$. The specific values chosen for D_1 and D_2 are the number of observations until December 2008, and 84 for the number of months from January 2009 to December 2015. α_{i0} , which governs the tightness of the prior distribution of P_{ii} , is set equal to 100, to ensure that regime switches do not occur too often.

- **Prior Distribution of $\mathcal{A}_1, \mathcal{A}_2$**

We noted earlier that $\mathcal{A}_{s_t}$ can be interpreted as the mean of the yield spreads given that the economy is in regime s_t . Therefore, we put the prior mean of $\mathcal{A}_1$ equal to the sample average of the yield spreads for the periods where the short rate is greater than 0.25, and the prior mean of $\mathcal{A}_2$ equal to the sample average of the yield spreads for the periods where the short rate is smaller than 0.25. If there are less than 24 observations for which the short rate is smaller than 0.25, then the prior mean of $\mathcal{A}_2$ computed in the above manner cannot be seen as representative of the unconditional mean of the yield spreads under the lower bound regime, so the prior means of both $\mathcal{A}_1$ and $\mathcal{A}_2$ are put equal to the sample average of the yield spreads across the entire sample.

- **Prior Distribution of $\sigma_{e,i}^2$**

We let $a_0^{(i)}, d_0^{(i)}$ be chosen so that the prior mean of $\sigma_{e,i}^2$ equals the sum of squared errors for the i th unobserved yield given that all the parameters are put equal to their prior means.

I MCMC Sampling Procedure

Here we detail our MCMC sampling procedure. Henceforth, $\pi(\cdot)$ indicates full conditional posterior densities, $l(\cdot)$ likelihoods and $p(\cdot)$ prior densities. Furthermore, the notation

$$\mathcal{N}(x \mid \mu, \Sigma)$$

refers to the value of the normal density with mean μ and covariance matrix Σ at the point x .

Following Hamilton and Wu (2012), we can reexpress the state-space model as a restricted Gaussian VAR. First, note that the factor dynamics are comprised of the regime-dependent level factor dynamics, and the regime-independent dynamics of the rest of the factors:

$$\begin{aligned} L_t &= \rho_{s_t} L_{t-1} + g_{s_t}^{\mathbb{P}_t} f_{t-1}^{SC} + \omega_{s_t} \cdot v_{1t}^{\mathbb{P}} \\ f_t^{SC} &= \tilde{G} f_{t-1}^{SC} + \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}}, \end{aligned}$$

where $v_{SC,t}^{\mathbb{P}}$ collects the last two elements of the vector of risk factors $v_t^{\mathbb{P}}$. Since f_t^{SC} is an affine function of the yield spreads $\mathcal{Y}\mathcal{S}_t^{(1)}$, we can see that

$$\begin{aligned} \mathcal{Y}\mathcal{S}_t^{(1)} &= \mathcal{A}_{s_t}^{(1)} + \mathcal{B}^{(1)} \left(\tilde{G} f_{t-1}^{SC} + \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}} \right) \\ &= \mathcal{A}_{s_t}^{(1)} + \mathcal{B}^{(1)} \tilde{G} \left(-\mathcal{B}^{(1)-1} \mathcal{A}_{s_{t-1}}^{(1)} + \mathcal{B}^{(1)-1} \mathcal{Y}\mathcal{S}_{t-1}^{(1)} \right) + \mathcal{B}^{(1)} \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}} \\ &= \left(\mathcal{A}_{s_t}^{(1)} - \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{A}_{s_{t-1}}^{(1)} \right) + \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{Y}\mathcal{S}_{t-1}^{(1)} + \mathcal{B}^{(1)} \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}} \end{aligned}$$

and

$$\begin{aligned}\mathcal{YS}_t^{(2)} &= \mathcal{A}_{s_t}^{(2)} + \mathcal{B}^{(2)} f_t^{SC} + \Sigma_e^{\frac{1}{2}} \cdot e_t \\ &= \left(\mathcal{A}_{s_t}^{(2)} - \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \mathcal{A}_{s_t}^{(1)} \right) + \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \cdot \mathcal{YS}_t^{(1)} + \Sigma_e^{\frac{1}{2}} \cdot e_t.\end{aligned}$$

Furthermore, identifying the level factor with the short rate and replacing f_{t-1}^{SC} with the affine function of $\mathcal{YS}_{t-1}^{(1)}$ allows us to reformulate the level factor dynamics as

$$r_t = -g_{s_t}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \mathcal{A}_{s_{t-1}}^{(1)} + \rho_{s_t} r_{t-1} + g_{s_t}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \cdot \mathcal{YS}_{t-1}^{(1)} + \omega_{s_t} \cdot v_{1t}^{\mathbb{P}}.$$

The three equations that characterize the model can be written as

$$r_t = \left(\bar{r}_{s_t} - g_{s_t}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \mathcal{A}_{s_{t-1}}^{(1)} \right) + \rho_{s_t} r_{t-1} + g_{s_t}^{\mathbb{P}'} \mathcal{B}^{(1)-1} \cdot \mathcal{YS}_{t-1}^{(1)} + \omega_{s_t} \cdot v_{1t}^{\mathbb{P}} \quad (\text{I.1})$$

$$\mathcal{YS}_t^{(1)} = \left(\mathcal{A}_{s_t}^{(1)} - \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{A}_{s_{t-1}}^{(1)} \right) + \mathcal{B}^{(1)} \tilde{G} \mathcal{B}^{(1)-1} \cdot \mathcal{YS}_{t-1}^{(1)} + \mathcal{B}^{(1)} \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}} \quad (\text{I.2})$$

$$\mathcal{YS}_t^{(2)} = \left(\mathcal{A}_{s_t}^{(2)} - \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \mathcal{A}_{s_t}^{(1)} \right) + \mathcal{B}^{(2)} \mathcal{B}^{(1)-1} \cdot \mathcal{YS}_t^{(1)} + \Sigma_e^{\frac{1}{2}} \cdot e_t. \quad (\text{I.3})$$

The parameters to be estimated are

$$\theta = \{\rho_2, g_1^{\mathbb{P}}, \omega_1^2, \omega_2^2, \tilde{G}, \Omega_{22}, \kappa^{\mathbb{Q}}, K_1^{\mathbb{Q}}, K_2^{\mathbb{Q}}, \Sigma_e, P\}$$

and the regime process $S = \{s_t\}_{1 \leq t \leq T}$.

Estimating the $\mathbb{Q}$ -parameter $K_{s_t}^{\mathbb{Q}}$ has traditionally proven difficult, not least because the separate estimation of $K_{s_t}^{\mathbb{Q}}$ and Ω_{s_t}, P means that the latter enter into the model in

a non-linear fashion. To circumvent this difficulty, we use the identity

$$\begin{pmatrix} K_1^{\mathbb{Q}} \\ K_2^{\mathbb{Q}} \end{pmatrix} = [c'_{11}c_{11} + c'_{12}c_{12}]^{-1} \left[c'_{11} \left(\mathcal{A}_1 + \frac{1}{2}c_{01} \right) + c'_{12} \left(\mathcal{A}_2 + \frac{1}{2}c_{02} \right) \right]. \quad (\text{I.4})$$

and estimate the parameters

$$\theta = \{\rho_2, g_1^{\mathbb{P}}, \omega_1^2, \omega_2^2, \tilde{G}, \Omega_{22}, \kappa^{\mathbb{Q}}, \mathcal{A}_1, \mathcal{A}_2, \Sigma_e, P\},$$

as mentioned in the main text.

Collect the observed variables as

$$X_t = \begin{pmatrix} r_t \\ \mathcal{Y}\mathcal{S}_t^{(1)} \\ \mathcal{Y}\mathcal{S}_t^{(2)} \end{pmatrix}.$$

We first discuss how to sample the parameters aside from $\kappa^{\mathbb{Q}}, \mathcal{A}_{s_t}$ and the regime S . The joint likelihood of the data X and the regime S is given as

$$\begin{aligned} l(X, S \mid \theta) &= \prod_{t=1}^T l(X_t, s_t \mid \mathcal{F}_{t-1}, \theta) \\ &= \prod_{t=1}^T l(X_t \mid s_t, \mathcal{F}_{t-1}, \theta) \cdot l(s_t \mid \mathcal{F}_{t-1}, \theta) \\ &= \prod_{t=1}^T l(X_t \mid s_t, \mathcal{F}_{t-1}, \theta) \cdot P_{s_{t-1}, s_t}, \end{aligned}$$

where $\mathcal{F}_t$ is the σ -algebra generated by the history of X_t and s_t up to time t . We can

further decompose each $l(X_t | s_t, \mathcal{F}_{t-1}, \theta)$ as

$$\begin{aligned}
l(X_t | s_t, \mathcal{F}_{t-1}, \theta) &= l(r_t, \mathcal{Y}\mathcal{S}_t^{(1)}, \mathcal{Y}\mathcal{S}_t^{(2)} | s_t, \mathcal{F}_{t-1}, \theta) \\
&= l(r_t | \mathcal{Y}\mathcal{S}_t, s_t, \mathcal{F}_{t-1}, \theta) \cdot l(\mathcal{Y}\mathcal{S}_t^{(2)} | \mathcal{Y}\mathcal{S}_t^{(1)}, s_t, \mathcal{F}_{t-1}, \theta) \cdot l(\mathcal{Y}\mathcal{S}_t^{(1)} | s_t, \mathcal{F}_{t-1}, \theta) \\
&= l(r_t | s_t, r_{t-1}, \mathcal{Y}\mathcal{S}_{t-1}^{(1)}; \{\kappa^{\mathbb{Q}}, \mathcal{A}_{s_t}^{(1)}, \rho_{s_t}, g_{s_t}^{\mathbb{P}}, \omega_{s_t}^2\}) \\
&\quad \times l(\mathcal{Y}\mathcal{S}_t^{(2)} | s_t, \mathcal{Y}\mathcal{S}_t^{(1)}; \{\kappa^{\mathbb{Q}}, \mathcal{A}_{s_t}, \Sigma_e\}) \\
&\quad \times l(\mathcal{Y}\mathcal{S}_t^{(1)} | s_t, \mathcal{Y}\mathcal{S}_{t-1}^{(1)}; \{\kappa^{\mathbb{Q}}, \mathcal{A}_{s_t}^{(1)}, \tilde{G}, \Omega_{22}\}).
\end{aligned}$$

Aside from $\kappa^{\mathbb{Q}}$ and $K^{\mathbb{Q}}$, each of the parameters appears only in one likelihood term, and thus only in one of the equations from (I.1) to (I.3). Since each of them are Gaussian regressions, we can either use the Normal-Normal or Inverse Gamma- Inverse Gamma (alternatively, Inverse Wishart- Inverse Wishart) update to recover the full conditional distribution of every one of them. We detail each step below:

Block 1: Sampling $\rho_{s_t}, g_{s_t}^{\mathbb{P}}, \omega_{s_t}^2$

The full conditional distribution of ρ_2 under the prior

$$\rho_2 \sim \mathcal{N}[\bar{\rho}_0, V_{\rho,0}]$$

is given as

$$\rho_2 | X, S, \theta \setminus \{\rho_2\} \sim \mathcal{N}[\bar{\rho}_1, V_{\rho,1}]$$

by the Normal-Normal update, where

$$V_{\rho,1} = \left[\frac{1}{V_{\rho,0}} + \frac{1}{\omega_2^2} \sum_{1 \leq t \leq T, s_t=2} r_{t-1}^2 \right]^{-1}$$

$$\bar{\rho}_1 = V_{\rho,1} \left[\frac{1}{\omega_2^2} \sum_{1 \leq t \leq T, s_t=2} r_t r_{t-1} + \frac{\bar{\rho}_0}{V_{\rho,0}} \right].$$

The full conditional distribution of $g_1^{\mathbb{P}}$ under the prior

$$g_1^{\mathbb{P}} \sim \mathcal{N}[\bar{g}_0, V_{g,0}]$$

is given as

$$g_1^{\mathbb{P}} \mid Y, S, \theta \setminus \{g_1^{\mathbb{P}}\} \sim \mathcal{N}[\bar{g}_1, V_{g,1}]$$

by the Normal-Normal update, where

$$V_{g,1} = \left[V_{g,0}^{-1} + \frac{1}{\omega_1^2} \sum_{1 \leq t \leq T, s_t=1} f_{t-1}^{SC} f_{t-1}^{SC'} \right]^{-1}$$

$$\bar{g}_1 = V_{g,1} \left[\frac{1}{\omega_1^2} \sum_{1 \leq t \leq T, s_t=1} f_{t-1}^{SC} \Delta r_t + V_{g,0}^{-1} \bar{g}_0 \right].$$

For any $i = 1, 2$, the full conditional distribution of ω_i^2 under the prior

$$\omega_i^2 \sim \mathcal{IG} \left[\frac{a_{i,0}}{2}, \frac{d_{i,0}}{2} \right]$$

is given as

$$\omega_i^2 \mid X, S, \theta \setminus \{\omega_i^2\} \sim \mathcal{IG} \left[\frac{a_{i,1}}{2}, \frac{d_{i,1}}{2} \right]$$

by the Inverse Gamma- Inverse Gamma update, where

$$\begin{aligned} a_{i,1} &= a_{i,0} + \#\{1 \leq t \leq T \mid s_t = i\} \\ d_{i,1} &= d_{i,0} + \sum_{1 \leq t \leq T, s_t = i} \left(r_t - \rho_i r_{t-1} - g_i^{\mathbb{P}'} f_{t-1}^{SC} \right)^2. \end{aligned}$$

Block 2: Sampling $\tilde{G}, \Omega_{22}$

Define $\gamma = \text{vec}(\tilde{G}^{\mathbb{P}'})$. The likelihood represented by equation (I.2) is equivalent to the likelihood of the equation

$$f_t^{SC} = \tilde{G} f_{t-1}^{SC} + \Omega_{22}^{\frac{1}{2}} \cdot v_{SC,t}^{\mathbb{P}}.$$

Thus, the full conditional distribution of γ under the prior

$$\gamma \sim \mathcal{N}[\bar{\gamma}_0, V_{\gamma,0}]$$

is given as

$$\gamma \mid X, S, \theta \setminus \{\gamma\} \sim \mathcal{N}[\bar{\gamma}_1, V_{\gamma,1}]$$

by the Normal-Normal update, where

$$V_{\gamma,1} = \left[V_{\gamma,0}^{-1} + \left(\Omega_{22}^{-1} \otimes \sum_{t=1}^T f_{t-1}^{SC} f_{t-1}^{SC'} \right) \right]^{-1}$$

$$\bar{\gamma}_1 = V_{\gamma,1} \left[\left(\Omega_{22}^{-1} \otimes I_2 \right) \cdot \text{vec} \left(\sum_{t=1}^T f_{t-1}^{SC} f_t^{SC'} \right) + V_{\gamma,0}^{-1} \bar{\gamma}_0 \right].$$

The full conditional distribution of Ω_{22} under the prior

$$\Omega_{22} \sim \mathcal{IW}[v_0, \Psi_0]$$

is given as

$$\Omega_{22} \mid X, S, \theta \setminus \{\Omega_{22}\} \sim \mathcal{IW}[v_1, \Psi_1]$$

by the Inverse Wishart- Inverse Wishart update, where

$$v_1 = v_0 + T$$

$$\Psi_1 = \Psi_0 + \sum_{t=1}^T \left(f_t^{SC} - \tilde{G}^{\mathbb{P}} f_{t-1}^{SC} \right) \left(f_t^{SC} - \tilde{G}^{\mathbb{P}} f_{t-1}^{SC} \right)'.$$

Block 3: Sampling Σ_e

The likelihood represented by equation (I.3) is equivalent to that represented

by

$$\mathcal{YS}_t^{(2)} = \mathcal{A}_{st}^{(2)} + \mathcal{B}^{(2)} f_t^{SC} + \Sigma_e^{\frac{1}{2}} \cdot e_t.$$

Therefore, for any $1 \leq i \leq m-2$, the full conditional distribution for $\sigma_{i,e}^2$ under its prior

$$\sigma_{i,e}^2 \sim \mathcal{IG} \left[\frac{a_0}{2}, \frac{d_0}{2} \right]$$

is given as

$$\sigma_{i,e}^2 \mid X, S, \theta \setminus \{\sigma_{i,e}^2\} \sim \mathcal{IG} \left[\frac{a_1}{2}, \frac{d_1}{2} \right]$$

under the Inverse Gamma- Inverse Gamme update, where

$$\begin{aligned} a_1 &= a_0 + T \\ d_1 &= d_0 + \sum_{t=1}^T \left(\mathcal{YS}_{it}^{(2)} - \mathcal{A}_{i,st}^{(2)} - \mathcal{B}_i^{(2)} f_t^{SC} \right)^2 \end{aligned}$$

and $\mathcal{YS}_{it}^{(2)}$, $\mathcal{A}_{i,st}^{(2)}$ and $\mathcal{B}_i^{(2)}$ collects the i th elements and i th row, respectively, of $\mathcal{YS}_t^{(2)}$, $\mathcal{A}_{st}^{(2)}$, and $\mathcal{B}^{(2)}$.

Block 4: Sampling P

For any $i = 1, 2$, letting j denote the regime that is not regime i , full conditional

distribution of P_{ii} under the prior

$$P_{ii} \sim \text{beta}[\alpha_{i,0}, \beta_{i,0}]$$

is given as

$$P_{ii} \mid X, S, \theta \setminus \{P_{ii}\} \sim \text{beta}[\alpha_{i,0} + n_{ii}, \beta_{i,0} + n_{ij}],$$

where n_{ii} is the number of times the regime changes from regime i to regime i and n_{ij} is the number of times the regime changes from regime i to regime j .

To estimate $\kappa^{\mathbb{Q}}$ and the regime process S , it is necessary to obtain a tractable expression for the likelihood of the data X . In addition, to estimate $\mathcal{A}_{s_t}$, we must collect the three equations characterizing the system together, since $\mathcal{A}_{s_t}^{(1)}$ appears in all three. To this end, we collect the three equations as

$$\begin{aligned} & \begin{pmatrix} 1 & O_{1 \times 2} & O_{1 \times (m-2)} \\ O_{2 \times 1} & I_2 & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & -\mathcal{B}^{(2)}\mathcal{B}^{(1)-1} & I_{m-2} \end{pmatrix} X_t \\ &= \begin{pmatrix} -g_{s_t}'\mathcal{B}^{(1)-1} & O_{1 \times (m-2)} \\ -\mathcal{B}^{(1)}\tilde{G}\mathcal{B}^{(1)-1} & O_{2 \times (m-2)} \\ O_{(m-2) \times 2} & O_{(m-2) \times (m-2)} \end{pmatrix} \underbrace{\begin{pmatrix} \mathcal{A}_{s_{t-1}}^{(1)} \\ \mathcal{A}_{s_{t-1}}^{(2)} \end{pmatrix}}_{\mathcal{A}_{s_{t-1}}} + \begin{pmatrix} O_{1 \times 2} & O_{1 \times (m-2)} \\ I_2 & O_{2 \times (m-2)} \\ -\mathcal{B}^{(2)}\mathcal{B}^{(1)-1} & I_{m-2} \end{pmatrix} \underbrace{\begin{pmatrix} \mathcal{A}_{s_t}^{(1)} \\ \mathcal{A}_{s_t}^{(2)} \end{pmatrix}}_{\mathcal{A}_{s_t}} \\ &+ \begin{pmatrix} \rho_{s_t} & g_{s_t}'\mathcal{B}^{(1)-1} & O_{1 \times (m-2)} \\ O_{2 \times 1} & \mathcal{B}^{(1)}\tilde{G}\mathcal{B}^{(1)-1} & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & O_{(m-2) \times 2} & O_{(m-2) \times (m-2)} \end{pmatrix} X_{t-1} + \begin{pmatrix} \omega_{s_t} & O_{1 \times 2} & O_{1 \times (m-2)} \\ O_{2 \times 1} & \mathcal{B}^{(1)}\Omega_{22}^{\frac{1}{2}} & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & O_{(m-2) \times 2} & \Sigma_{\tilde{e}}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} v_{1t}^{\mathbb{P}} \\ v_{SC,t}^{\mathbb{P}} \\ e_t \end{pmatrix}. \end{aligned}$$

Since

$$\begin{pmatrix} 1 & O_{1 \times 2} & O_{1 \times (m-2)} \\ O_{2 \times 1} & I_2 & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & -\mathcal{B}^{(2)}\mathcal{B}^{(1)-1} & I_{m-2} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & O_{1 \times 2} & O_{1 \times (m-2)} \\ O_{2 \times 1} & I_2 & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & \mathcal{B}^{(2)}\mathcal{B}^{(1)-1} & I_{m-2} \end{pmatrix},$$

defining

$$D_{1,s_t} = \begin{pmatrix} -g_{s_t}^{\mathbb{P}'}\mathcal{B}^{(1)-1} & O_{1 \times (m-2)} \\ -\mathcal{B}^{(1)}\tilde{G}\mathcal{B}^{(1)-1} & O_{2 \times (m-2)} \\ -\mathcal{B}^{(2)}\tilde{G}\mathcal{B}^{(1)-1} & O_{(m-2) \times (m-2)} \end{pmatrix}$$

$$E = \begin{pmatrix} O_{1 \times m} \\ I_m \end{pmatrix}$$

$$C_{1,s_t,s_{t-1}} = \begin{pmatrix} D_{1,s_t} \cdot (2 - s_{t-1}) & D_{1,s_t} \cdot (s_{t-1} - 1) \end{pmatrix} + \begin{pmatrix} E \cdot (2 - s_t) & E \cdot (s_t - 1) \end{pmatrix}$$

the reduced form of this system is

$$X_t = C_{1,s_t,s_{t-1}} \cdot \underbrace{\begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{pmatrix}}_{\mathcal{A}} + \underbrace{\begin{pmatrix} \rho_{s_t} & g_{s_t}^{\mathbb{P}'}\mathcal{B}^{(1)-1} & O_{1 \times (m-2)} \\ O_{2 \times 1} & \mathcal{B}^{(1)}\tilde{G}\mathcal{B}^{(1)-1} & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & \mathcal{B}^{(2)}\tilde{G}\mathcal{B}^{(1)-1} & O_{(m-2) \times (m-2)} \end{pmatrix}}_{\Phi_{s_t}} X_{t-1} + \underbrace{\begin{pmatrix} \omega_{s_t} & O_{1 \times 2} & O_{1 \times (m-2)} \\ O_{2 \times 1} & \mathcal{B}^{(1)}\Omega_{22}^{\frac{1}{2}} & O_{2 \times (m-2)} \\ O_{(m-2) \times 1} & \mathcal{B}^{(2)}\Omega_{22}^{\frac{1}{2}} & \Sigma_{\tilde{e}}^{\frac{1}{2}} \end{pmatrix}}_{\Sigma_{s_t}^{\frac{1}{2}}} \underbrace{\begin{pmatrix} v_{1t}^{\mathbb{P}} \\ v_{SC,t}^{\mathbb{P}} \\ e_t \end{pmatrix}}_{u_t}.$$

The time t likelihood given the time t and $t - 1$ regimes now becomes

$$l(X_t | s_t, s_{t-1}, X_{t-1}, \theta) = \left(\frac{1}{2\pi}\right)^{\frac{m+1}{2}} |\Sigma_{s_t}|^{-\frac{1}{2}} \\ \times \exp\left(-\frac{1}{2} \left(X_t - C_{1,s_t,s_{t-1}}\mathcal{A} - \Phi_{s_t}X_{t-1}\right)' \Sigma_{s_t}^{-1} \left(X_t - C_{1,s_t,s_{t-1}}\mathcal{A} - \Phi_{s_t}X_{t-1}\right)\right),$$

and

$$l(X, S | \theta) = \prod_{t=1}^T l(X_t | s_t, s_{t-1}, X_{t-1}, \theta) \cdot P_{s_{t-1}, s_t} \\ = \left(\frac{1}{2\pi}\right)^{\frac{T(m+1)}{2}} \left| \prod_{t=1}^T \Sigma_{s_t} \right|^{-\frac{1}{2}} \cdot \left(\prod_{t=1}^T P_{s_{t-1}, s_t} \right) \\ \times \exp\left(-\frac{1}{2} \sum_{t=1}^T \left(X_t - C_{1,s_t,s_{t-1}}\mathcal{A} - \Phi_{s_t}X_{t-1}\right)' \Sigma_{s_t}^{-1} \left(X_t - C_{1,s_t,s_{t-1}}\mathcal{A} - \Phi_{s_t}X_{t-1}\right)\right).$$

It follows that $\kappa^{\mathbb{Q}}, \mathcal{A}_{s_t}$ and the regime process S can be sampled as follows:

Block 5: Sampling $\kappa^{\mathbb{Q}}$ and $\mathcal{A}$

Since $\kappa^{\mathbb{Q}}$ is assumed to take finitely many values, we can just take the probability that $\kappa^{\mathbb{Q}}$ equals a certain value as proportional to the likelihood value under this value.

As for $\mathcal{A}$, we can use the fact that it serves as the intercept term in the regression

$$X_t = C_{1,s_t,s_{t-1}}\mathcal{A} + \Phi_{s_t}X_{t-1} + \Sigma_{s_t}^{\frac{1}{2}} \cdot u_t.$$

It follows that the full conditional distribution of $\mathcal{A}$ under the prior

$$\mathcal{A} \sim \mathcal{N}[\bar{\mathcal{A}}_0, V_{\mathcal{A},0}]$$

is given as

$$\mathcal{A} \mid X, S, \theta \setminus \{\mathcal{A}\} \sim \mathcal{N}[\bar{\mathcal{A}}_1, V_{\mathcal{A},1}]$$

by the Normal-Normal update, where

$$V_{\mathcal{A},1} = \left[\sum_{t=1}^T C'_{1,s_t,s_{t-1}} \Sigma_{s_t}^{-1} C_{1,s_t,s_{t-1}} + V_{\mathcal{A},0}^{-1} \right]^{-1}$$

$$\bar{\mathcal{A}}_{i,1} = V_{\mathcal{A},1} \left[\sum_{t=1}^T C'_{1,s_t,s_{t-1}} \Sigma_{s_t}^{-1} (X_t - \Phi_{s_t} X_{t-1}) + V_{\mathcal{A},0}^{-1} \bar{\mathcal{A}}_0 \right].$$

Block 6: Sampling S

It now remains to sample the regime process. This is done through the standard Hamilton filter combined with the Carter-Kohn backward recursion. By assumption, information on whether the economy is at the zero lower bound at a certain point in time is conveyed through the factors,

Let $\mathbf{X}_t$ be the σ -algebra generated by the history of X_t up to time t , and define the quantities

$$\alpha_{t|s} = \begin{pmatrix} \mathbb{P}(s_t = 1 \mid \mathbf{X}_s) \\ \vdots \\ \mathbb{P}(s_t = N \mid \mathbf{X}_s) \end{pmatrix}$$

for any $0 \leq t, s \leq T$. Then, the predictive and filtered probabilities can be com-

puted recursively, and the regime process sampled, as follows:

1) **Step 0: Loading Initial Values**

Prepare initial values by setting

$$\alpha_{0|0} = \begin{pmatrix} \frac{1-P_{22}}{2-P_{11}-P_{22}} \\ \frac{1-P_{11}}{2-P_{11}-P_{22}} \end{pmatrix},$$

the stationary distribution of the Markov chain.

2) **Step 1: Predictive Probabilities**

Given $\alpha_{t-1|t-1}$, we can compute

$$\alpha_{t|t-1} = P' \cdot \alpha_{t-1|t-1}.$$

3) **Step 2: Filtered Probabilities**

For any $1 \leq i \leq N$,

$$\begin{aligned} \mathbb{P}(s_t = i, s_{t-1} = j \mid \mathbf{X}_t) &\propto l(X_t \mid s_t = i, s_{t-1} = j, X_{t-1}) \cdot \mathbb{P}(s_t = i, s_{t-1} = j \mid \mathbf{X}_{t-1}) \\ &= l(X_t \mid s_t = i, s_{t-1} = j, X_{t-1}) \cdot \mathbb{P}(s_t = i \mid s_{t-1} = j) \cdot \mathbb{P}(s_{t-1} = j \mid \mathbf{X}_{t-1}) \\ &= l(X_t \mid s_t = i, s_{t-1} = j, X_{t-1}) \cdot P_{ji} \cdot \alpha_{t-1|t-1,j}. \end{aligned}$$

Since

$$\mathbb{P}(s_t = i \mid \mathbf{X}_t) = \sum_{j=1}^N \mathbb{P}(s_t = i, s_{t-1} = j \mid \mathbf{X}_t),$$

it follows that the kernels of the filtered probabilities in $\alpha_{t|t}$ are given as

$$k_{t|t} = \begin{pmatrix} \sum_{j=1}^N l(X_t | s_t = 1, s_{t-1} = j, X_{t-1}) \cdot P_{j1} \cdot \alpha_{t-1|t-1,j} \\ \vdots \\ \sum_{j=1}^N l(X_t | s_t = N, s_{t-1} = j, X_{t-1}) \cdot P_{jN} \cdot \alpha_{t-1|t-1,j} \end{pmatrix},$$

and the filtered probabilities as

$$\alpha_{t|t} = \frac{1}{\iota' k_{t|t}} k_{t|t}.$$

4) **Step 3: Sampling Regimes**

Sample s_T according to the filtered probabilities $\alpha_{T|T}$. Having sampled s_{t+1} , calculate the probabilities $\alpha_{t|T}$ as

$$\alpha_{t|T} = \frac{1}{\sum_{i=1}^N P_{i,s_{t+1}} \cdot \alpha_{t|t,i}} \begin{pmatrix} P_{1,s_{t+1}} \cdot \alpha_{t|t,1} \\ \vdots \\ P_{N,s_{t+1}} \cdot \alpha_{t|t,N} \end{pmatrix}$$

s_t is then sampled according to the probabilities $\alpha_{t|T}$.

Given the sample $\{\theta^{(i)}, S^{(i)}\}$ of the parameters and regime drawn from the posterior distribution $\theta, S | \mathbf{X}$, we compute point estimates of quantities such as the term premium and the future short rate as follows. Suppose we want to estimate the function $f(\mathbf{X}, \theta, S)$ of both the data and parameters. Under the quadratic loss function

$$L(\hat{f} | \mathbf{X}) = \mathbb{E}_{\theta, S} \left[\left| f(\mathbf{X}, \theta, S) - \hat{f} \right|^2 | \mathbf{X} \right]$$

$$= \int_{\Theta \times \mathcal{S}} |f(\mathbf{X}, \theta, S) - \hat{f}|^2 \pi(\theta, S | \mathbf{X}) d(\theta, S),$$

it is known that the optimal point estimator of $f(\mathbf{X}, \theta, S)$ is the expected value

$$\mathbb{E}_{\theta, S}[f(\mathbf{X}, \theta, S) | \mathbf{X}] = \int_{\Theta \times \mathcal{S}} f(\mathbf{X}, \theta, S) \pi(\theta, S | \mathbf{X}) d(\theta, S).$$

Since we have a sample $\{\theta^{(i)}, S^{(i)}\}$ drawn from the posterior distribution, the above quantity can be computed as

$$\hat{f}(\mathbf{X}) = \frac{1}{N_1} \sum_{i=1}^{N_1} f(\mathbf{X}, \theta^{(i)}, S^{(i)}),$$

where N_1 is the size of the sample.

J Computing the Term Premium

The term premium is given as the average expected excess return over a given horizon, which are functions of regimes and factor processes, so it be computationally costly to obtain through numerical methods, especially for longer maturities. Here we present a simple algorithm for the calculation of the term premium that greatly reduces the computational cost.

Suppose that the regime process S and the sampled values of Ω_{s_t} , $G_{s_t}^{\mathbb{P}}$, $\kappa^{\mathbb{Q}}$ and P are given. The τ -maturity term premium at time t is a function of these parameters and the time t regime:

$$\begin{aligned} TP_t(\tau) &= \frac{1}{\tau} \sum_{i=0}^{\tau-2} \mathbb{E}_t \left[-\frac{1}{2} b(\tau-i-1)' \Omega_{s_{t+i+1}} b(\tau-i-1) + b(\tau-i-1)' \kappa^{\mathbb{Q}} \mid s_t \right] \\ &\quad + \frac{1}{\tau} \sum_{i=0}^{\tau-2} \mathbb{E}_t \left[-b(\tau-i-1)' (G_{s_{t+i+1}}^{\mathbb{P}} - G^{\mathbb{Q}}) f_{t+i} \mid s_t \right]. \end{aligned}$$

Exploiting the Markov property of the regimes, for any $0 \leq i \leq \tau - 2$

$$\mathbb{E}_t \left[\Omega_{s_{t+i+1}} \mid s_t \right] = \mathbb{E} \left[\Omega_{s_{i+1}} \mid s_0 \right].$$

Because $\mathbb{P}(s_{i+1} = j \mid s_0) = P_{s_0, j}^{i+1}$ for $j = 1, 2$,

$$\mathbb{E}_t \left[\Omega_{s_{t+i+1}} \mid s_t \right] = \sum_{j=1}^2 P_{s_0, j}^{i+1} \cdot \Omega_j.$$

The tricky part of estimating the term premium lies with the factor component $(G_{s_{t+i+1}}^{\mathbb{P}} - G^{\mathbb{Q}}) \cdot f_{t+i}$, since it is the product of two random elements. Below we show how to recursively compute the expectation of the factor component.

Note initially that

$$\mathbb{E}_t \left[G_{s_{t+1}}^{\mathbb{P}} f_t \mid s_t \right] = \mathbb{E} \left[G_{s_{t+1}}^{\mathbb{P}} \mid s_t \right] \cdot f_t.$$

Suppose now that, for some $k \geq 1$,

$$\mathbb{E}_t \left[G_{s_{t+k}}^{\mathbb{P}} f_{t+k-1} \mid s_t \right] = \mathbb{E} \left[\prod_{h=1}^k G_{s_{t+h}}^{\mathbb{P}} \mid s_t \right] \cdot f_t$$

for any t . Then, because the factors follow the VAR(1) process

$$f_{t+1} = G_{s_{t+1}}^{\mathbb{P}} f_t + \varepsilon_{t+1}^{\mathbb{P}},$$

we have, by the law of iterated expectations,

$$\mathbb{E}_t \left[G_{s_{t+k+1}}^{\mathbb{P}} f_{t+k} \mid s_{t+1}, s_t \right] = \mathbb{E}_t \left[\mathbb{E}_{t+1} \left[G_{s_{t+k+1}}^{\mathbb{P}} f_{t+k} \mid s_{t+1} \right] \mid s_{t+1}, s_t \right]$$

$$\begin{aligned}
&= \mathbb{E} \left[\prod_{h=2}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} \right] \cdot \mathbb{E}_t[f_{t+1} \mid s_{t+1}, s_t] \\
&\quad \text{(Inductive Hypothesis)} \\
&= \mathbb{E} \left[\prod_{h=2}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} \right] \cdot \mathbb{E}_t \left[G_{s_{t+1}}^{\mathbb{P}} \cdot f_t + \varepsilon_{t+1}^{\mathbb{P}} \mid s_{t+1}, s_t \right] \\
&= \mathbb{E} \left[\prod_{h=2}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} \right] \cdot \left(G_{s_{t+1}}^{\mathbb{P}} \cdot f_t \right) \\
&= \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} \right] \cdot f_t.
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_t \right] &= \sum_{j=1}^2 \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} = j, s_t \right] \cdot P_{s_t, j} \\
&= \sum_{j=1}^2 \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} = j \right] \cdot P_{s_t, j},
\end{aligned}$$

where the last equality follows by the Markov property, it follows that

$$\begin{aligned}
\mathbb{E}_t \left[G_{s_{t+k+1}}^{\mathbb{P}} f_{t+k} \mid s_t \right] &= \sum_{j=1}^2 \mathbb{E}_t \left[G_{s_{t+k+1}}^{\mathbb{P}} f_{t+k} \mid s_{t+1} = j, s_t \right] \cdot P_{s_t, j} \\
&= \left[\sum_{j=1}^2 \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_{t+1} = j \right] \cdot P_{s_t, j} \right] \cdot f_t \\
&= \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_t \right] \cdot f_t.
\end{aligned}$$

By induction,

$$\mathbb{E}_t \left[G_{s_{t+k+1}}^{\mathbb{P}} f_{t+k} \mid s_t \right] = \mathbb{E} \left[\prod_{h=1}^{k+1} G_{s_{t+h}}^{\mathbb{P}} \mid s_t \right] \cdot f_t$$

for any $k \geq 1$, and similarly, we have

$$\mathbb{E}_t[f_{t+k} \mid s_t] = \mathbb{E} \left[\prod_{h=1}^k G_{s_{t+h}}^{\mathbb{P}} \mid s_t \right] \cdot f_t$$

for any $k \geq 1$. Furthermore, due to the memoryless property of the Markov chain $\{s_t\}$,

for any $k \geq 1$ and $j = 1, 2$,

$$\mathbb{E} \left[\prod_{h=1}^k G_{s_{t+h}}^{\mathbb{P}} \mid s_t = j \right] = \mathbb{E} \left[\prod_{h=1}^k G_{s_h}^{\mathbb{P}} \mid s_0 = j \right] = EG_k(j),$$

so that the expectation depends only on the “initial” value s_t .

The factor component of the term premium can now be expressed as

$$\begin{aligned} \frac{1}{\tau} \sum_{i=0}^{\tau-2} \mathbb{E}_t \left[-b(\tau-i-1)' (G_{s_{t+i+1}}^{\mathbb{P}} - G^{\mathbb{Q}}) f_{t+i} \mid s_t \right] \\ = \frac{1}{\tau} \sum_{i=0}^{\tau-2} \left[-b(\tau-i-1)' (EG_{i+1}(s_t) - G^{\mathbb{Q}} \cdot EG_i(s_t)) f_t \right]. \end{aligned}$$

Therefore, we need only know how to compute the expectations $EG_i(j)$ for $1 \leq i \leq \tau-1$ and $j = 1, 2$ to be able to calculate the term premium.

To do so, we proceed by induction. It is easy to see that

$$EG_1(j) = \mathbb{E} \left[G_{s_1}^{\mathbb{P}} \mid s_0 = j \right] = \sum_{k=1}^2 G_k^{\mathbb{P}} \cdot P_{jk}$$

for $j = 1, 2$. Now suppose that we know $EG_i(1)$ and $EG_i(2)$ for some $1 \leq i \leq \tau-2$. Then, by the law of iterated expectations, for $j = 1, 2$

$$\begin{aligned} EG_{i+1}(j) &= \mathbb{E} \left[G_{s_{i+1}}^{\mathbb{P}} \times \cdots \times G_{s_1}^{\mathbb{P}} \mid s_0 = j \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[G_{s_{i+1}}^{\mathbb{P}} \times \cdots \times G_{s_2}^{\mathbb{P}} \mid s_1, s_0 = j \right] \cdot G_{s_1}^{\mathbb{P}} \mid s_0 = j \right] \\ &= \mathbb{E} \left[EG_i(s_1) \cdot G_{s_1}^{\mathbb{P}} \mid s_0 = j \right] \end{aligned}$$

$$= \sum_{k=1}^2 EG_i(k) G_k^{\mathbb{P}} \cdot P_{jk}.$$

where we utilized the Markov property of the regime process to justify the third equality.

Using this recursion, we can sequentially compute the values of $EG_1(\cdot), \dots, EG_{\tau-1}(\cdot)$ starting from $EG_1(\cdot)$.

In summary, in order to compute $TP_t(\tau)$ for any time t and maturity τ , we proceed as follows:

1. Initial Values of EG Terms

First compute

$$EG_1(j) = \sum_{k=1}^2 G_k^{\mathbb{P}} \cdot P_{jk}$$

for $j = 1, 2$.

2. Recursive Computation of EG Terms

For any $1 \leq i \leq \tau - 2$ and $j = 1, 2$, compute

$$EG_{i+1}(j) = \sum_{k=1}^2 EG_i(k) G_k^{\mathbb{P}} \cdot P_{jk}.$$

3. Obtaining $TP_t(\tau)$

Finally, using the EG terms obtained above, compute

$$\begin{aligned} TP_t(\tau) = & \frac{1}{\tau} \sum_{i=0}^{\tau-2} \left[-\frac{1}{2} b(\tau-i-1)' \left(\sum_{j=1}^2 \Omega_j \cdot P_{st,j}^{i+1} \right) b(\tau-i-1) + b(\tau-i-1)' \mathbb{Q} \right] \\ & + \frac{1}{\tau} \sum_{i=0}^{\tau-2} \left[-b(\tau-i-1)' \left(EG_{i+1}(s_t) - G^{\mathbb{Q}} \cdot EG_i(s_t) \right) f_t \right]. \end{aligned}$$

K Recursive Expansion of Factors

Here we present the recursive expansion of the h -period ahead factors f_{t+h} and how it can be used to compute conditional yield volatilities.

The physical factor dynamics tell us that

$$f_{t+h} = G_{s_{t+h}}^{\mathbb{P}} f_{t+h-1} + \Omega_{s_{t+h}}^{\frac{1}{2}} \cdot v_{t+h}^{\mathbb{P}},$$

where $v_{t+3}^{\mathbb{P}} \sim \mathcal{N}[O_{3 \times 1}, I_3]$ defines a conditionally independent innovation process. Expanding this recursion until we obtain the time t factors f_t , we have

$$f_{t+h} = \left(\prod_{i=1}^h G_{s_{t+i}}^{\mathbb{P}} \right) f_t + \Omega_{s_{t+h}}^{\frac{1}{2}} \cdot v_{t+h}^{\mathbb{P}} + \sum_{j=1}^{h-1} \left(\prod_{i=j+1}^h G_{s_{t+i}}^{\mathbb{P}} \right) \Omega_{s_{t+j}}^{\frac{1}{2}} \cdot v_{t+j}^{\mathbb{P}},$$

where we let

$$\prod_{i=j+1}^h G_{s_{t+i}}^{\mathbb{P}} = G_{s_{t+h}}^{\mathbb{P}} \times \cdots \times G_{s_{t+j+1}}^{\mathbb{P}}$$

for any $0 \leq j \leq h-1$.

Taking time t expectations, given the regimes $s_{t+1}, \dots, s_{t+h}$ up to time $t+h$, on both sides of the recursion yields

$$\mathbb{E}_t[f_{t+h} \mid s_{t+1}, \dots, s_{t+h}] = \left(\prod_{i=1}^h G_{s_{t+i}}^{\mathbb{P}} \right) f_t,$$

since

$$\mathbb{E}_t[v_{t+j}^{\mathbb{P}} \mid s_{t+1}, \dots, s_{t+h}] = O_{3 \times 1}$$

by the idiosyncratic nature of the errors.

Similarly,

$$\begin{aligned}
& \text{Var}_t [f_{t+h} \mid s_{t+1}, \dots, s_{t+h}] \\
&= \mathbb{E}_t \left[(f_{t+h} - \mathbb{E}_t [f_{t+h} \mid s_{t+1}, \dots, s_{t+h}]) (f_{t+h} - \mathbb{E}_t [f_{t+h} \mid s_{t+1}, \dots, s_{t+h}])' \mid s_{t+1}, \dots, s_{t+h} \right] \\
&= \text{Var}_t \left[\Omega_{s_{t+h}}^{\frac{1}{2}} \cdot v_{t+h}^{\mathbb{P}} + \sum_{j=1}^{h-1} \left(\prod_{i=j+1}^h G_{s_{t+i}}^{\mathbb{P}} \right) \Omega_{s_{t+j}}^{\frac{1}{2}} \cdot v_{t+j}^{\mathbb{P}} \mid s_{t+1}, \dots, s_{t+h} \right] \\
&= \Omega_{s_{t+h}} + \sum_{j=1}^{h-1} \left(\prod_{i=j+1}^h G_{s_{t+i}}^{\mathbb{P}} \right) \Omega_{s_{t+j}} \left(\prod_{i=j+1}^h G_{s_{t+i}}^{\mathbb{P}} \right)'.
\end{aligned}$$

We show how this expansion can be used to compute conditional yield volatilities.

The model-implied yield of a τ -maturity bond is

$$Y_t(\tau) = \frac{a_{s_t}(\tau)}{\tau} + \frac{b(\tau)'}{\tau} f_t.$$

As above, the h -horizon ahead conditional volatility for this yield given that the economy is at regime j now and for the next h periods is

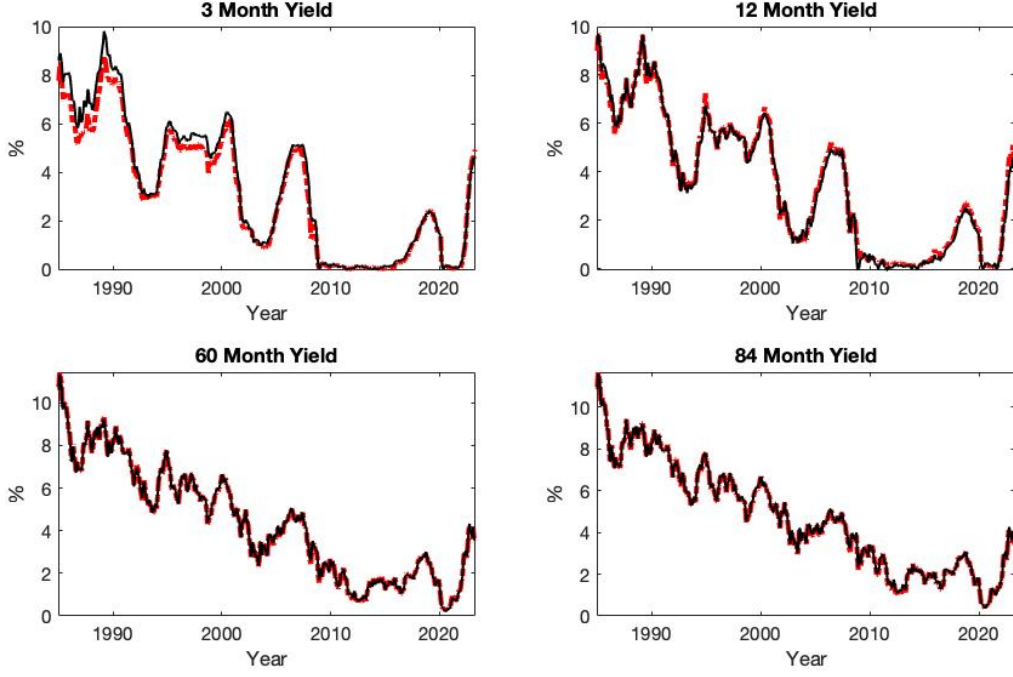
$$\begin{aligned}
\text{Var}_t [Y_{t+h}(\tau) \mid s_t = j, s_{t+1} = j, \dots, s_{t+h} = j] &= \frac{1}{\tau^2} b(\tau)' \text{Var}_t [f_{t+h} \mid s_t = j, s_{t+1} = j, \dots, s_{t+h} = j] b(\tau) \\
&= \frac{1}{\tau^2} b(\tau)' \left[\Omega_j + \sum_{k=1}^{h-1} \left(G_j^{\mathbb{P}} \right)^{h-k} \Omega_j \left(G_j^{\mathbb{P}'} \right)^{h-k} \right] b(\tau).
\end{aligned}$$

Here we assume that the regimes for the next h periods is fixed because integrating out the regimes $s_{t+1}, \dots, s_{t+h}$ is complicated, as it requires using the law of conditional variances.

L Additional Results Related to Model Fit

Here we present some empirical results related to model fit, with an emphasis on the ability of our model to capture changes in yield dynamics related between the normal

Figure 1: **Fitted and Actual Yields** This figure plots the yields and their fitted values for selected maturities. The dashed lines are the observed yields, and the solid lines are the fitted values.



and ZLB regime.

L.1 Yield Fit

We obtain a sample of the model-implied τ maturity yield by computing

$$\frac{a_{s_t}(\tau)}{\tau} + \frac{b(\tau)'}{\tau} f_t,$$

using the values of the parameters procured in each iteration of the Gibbs sampling procedure. Here, $a_{s_t}(\tau)$ is given as the function linear in $K_1^{\mathbb{Q}}, K_2^{\mathbb{Q}}$ and non-linear in

Ω_1, Ω_2, P , as per Equation (A.1). f_t , meanwhile, is recovered from the equation

$$\mathcal{Y}\mathcal{S}_t^{(1)} = \mathcal{A}_{s_t}^{(1)} + \mathcal{B}^{(1)} f_t$$

as

$$f_t = \mathcal{B}^{(1)-1} \left[\mathcal{Y}\mathcal{S}_t^{(1)} - \mathcal{A}_{s_t}^{(1)} \right],$$

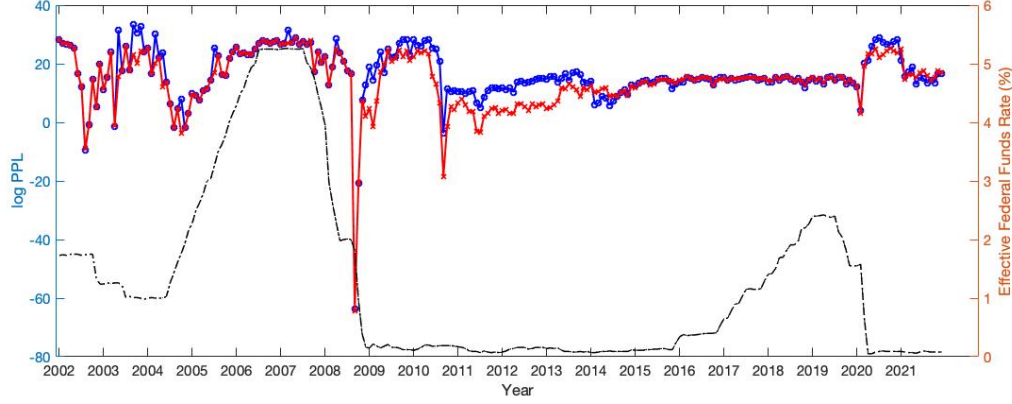
where $\mathcal{A}_{s_t}$ here is directly sampled in each iteration and does not necessarily satisfy the no-arbitrage restrictions in Equation (A.1). Nevertheless, as emphasized in the previous section, the estimator of $\mathcal{A}_{s_t}$ asymptotically satisfies these restrictions.

Figure 1 presents the actual yields alongside their model-implied counterparts for maturities of three months, one year, two years, and seven years. Because the effective federal funds rate serves as our proxy for the one-month rate, the measurement errors for shorter maturity yields are quite noticeable when compared with those of longer maturities. The most significant discrepancies between the actual yields and the fitted yields are found earlier in the sample, well before the economy hits the ZLB. One possible interpretation for the overestimation of the 3 month yield during the earlier periods of the sample is that the imposition of the ZLB restriction, which prevents yields from becoming highly negative, has the side effect of driving the model-implied yields upward.

L.2 Overall Forecasting Ability

In the preceding section, we showed how incorporating the ZLB into the FS-only model via regime switching improves short rate forecasts by imposing the restriction that the short rate cannot become highly negative. Now we compare how the FS-ZLB model fares against the FS-only model when it comes to overall forecasting ability. Instead

Figure 2: **Log PPLs** This figure plots log PPLs of the FS-only and FS-ZLB models during the out-of-sample period from January 2002 to December 2021. The round markers represent the forecasts of the FS-ZLB model, and the crossed markers those of the FS-only model. The dotted-dashed line is the effective federal funds rate.



of conducting out-of-sample forecasting exercises for each individual yield, we rely on the posterior predictive likelihood (PPL) as our measure of overall forecasting ability. Given a sample until time T and data at time $T+1$, the PPL is the one-period ahead likelihood $l(X_{T+1} | X_T, \dots, X_1)$, where X_t is defined as in expression (H.6). Heuristically, it represents how likely X_{T+1} is to be observed given the past data; by implication, the greater the PPL, the more likely X_{T+1} would have been predicted given the past data, which allows it to serve as a measure of forecasting ability.

Given a sample $\{\theta^{(i)}, S^{(i)}\}_{1 \leq i \leq N}$ of size N Gibbs sampled from the posterior distribution $\theta, S | X_T, \dots, X_1$, the PPL is estimated as

$$\frac{1}{N} \sum_{i=1}^N l(X_{T+1} | X_T, \dots, X_1, \theta^{(i)}, S^{(i)}),$$

where the likelihood $l(X_{T+1} | X_T, \dots, X_1, \theta^{(i)}, S^{(i)})$ at each iteration is computed based on Equation (H.6).

Figure 2 displays the log PPLs obtained from the FS-ZLB model against those of the FS-only model. It is immediately evident that the PPLs obtained from the FS-ZLB

Table 1: **Estimated Regime-Switching Parameters** Tabulated below are the estimated level factor parameters, taken as the mean of the sampled values. Standard errors are in parentheses.

	Level Factor Parameters			
	ρ_{st}	$g_{1,st}^{\mathbb{P}}$	$g_{2,st}^{\mathbb{P}}$	ω_{st}^2
Normal Regime	1 (0)	0.048*** (0.0063)	0.073*** (0.005)	0.032*** (0.002)
Lower Bound Regime	0.814*** (0.01)	0 (0)	0 (0)	0.001*** (0.01)

Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

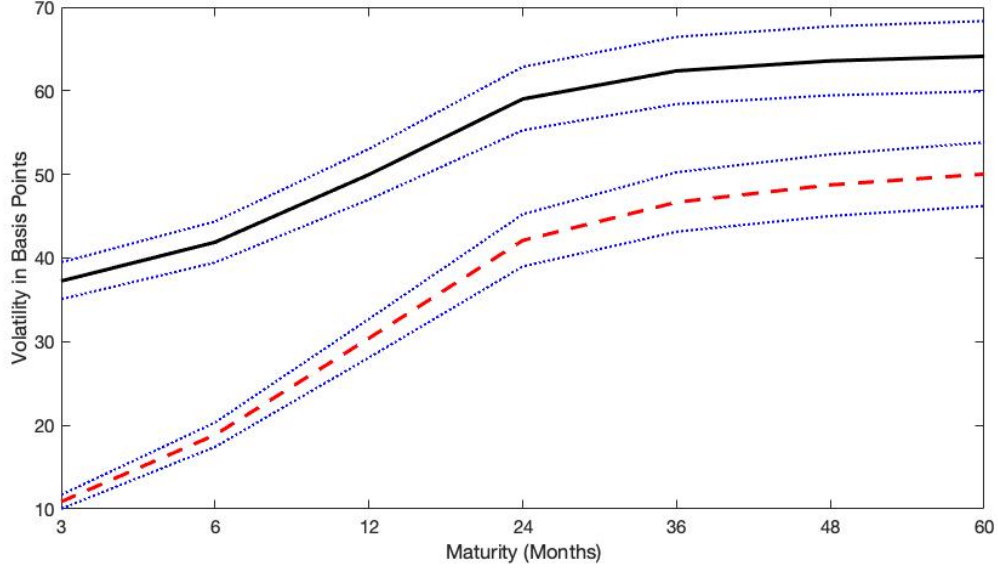
model largely coincide with those of the FS-only model, just like how the two models reported similar short rate forecasts for a large portion of the sample. As in the case of the short rate forecasts, the greatest point of divergence between the two models is during the lower bound regime, most notably during the periods belonging to the GFC. This implies that adherence to the ZLB restriction not only improves forecasts of the short rate, but also benefits forecasting ability in general.

L.2.1 Regime-Switching Parameters

In Table 1, we present the estimated values of the regime-switching level factor parameters ρ_{st} , $g_{st}^{\mathbb{P}}$ and ω_{st}^2 . Both elements of $g_1^{\mathbb{P}}$ are positive, which tells us that, during the normal regime, the short rate is positively correlated with the previous period's slope and curvature factors. As will be explained in more detail below, this effectively adds the variation of the slope and curvature factors to the variation of the level factor, so that, during the normal regime, the level factor becomes much more volatile than what the estimated value of ω_1^2 implies.

On the other hand, the level factor is very clearly a stationary AR(1) process during the lower bound regime, since it only depends on the past value of the level factor. As we

Figure 3: **Conditional Yield Volatilities** This figure plots the estimated 3-month ahead conditional yield volatilities for selected maturities and the corresponding 95% confidence bands. The solid line represents the volatilities in the normal regime, and the dashed line is those of the lower bound regime.



intended, the error variance of the level factor is much higher during the normal regime compared to the lower bound regime, with ω_1 being on average almost 7 times as higher than ω_2 .

L.2.2 Conditional Yield Volatility

To further assess whether our model effectively differentiates between the normal and lower-bound regimes, we examine the estimated regime-specific conditional yield volatilities. According to Swanson and J. C. Williams (2014), the volatility of short-term yields are compressed during the ZLB, the effects of which extend even to longer-term yields. Here we present evidence that our model accurately captures this effect.

As is customary in the literature, we investigate conditional yield volatilities (e.g., Christensen and Rudebusch (2016)). The h -period-ahead conditional volatility of a τ -maturity yield at time t , where the regimes at times $t+1$ to $t+h$ are known, is defined

as

$$\text{Var}_t[Y_{t+h}(\tau) \mid s_{t+1}, \dots, s_{t+h}] = \frac{1}{\tau^2} b(\tau)' \text{Var}_t[f_{t+h} \mid s_{t+1}, \dots, s_{t+h}] b(\tau).$$

The conditional covariance of f_{t+h} is computed as in Appendix K. We estimate the three-month-ahead conditional volatilities, as in Christensen and Rudebusch (2016), for the two cases $s_t = \dots = s_{t+3} = 1$ (for regime 1) and $s_t = \dots = s_{t+3} = 2$ (for regime 2).

The estimated volatilities and their respective 95% confidence intervals are shown in Figure 3. As anticipated, yield volatilities are higher in the normal regime than in the lower bound regime. In particular, the conditional yield volatility for three-month yields is approximately 10 basis points, aligning closely with the realized volatilities for the three-month yield reported by Christensen and Rudebusch (2016).

Our model also proves capable of replicating the finding in Swanson and J. C. Williams (2014) that short-term yields are more significantly affected by the lower bound than longer-term yields. Specifically, the conditional yield volatility for three-month yields in the lower bound regime is approximately 25 to 30 basis points lower than in the normal regime, while, for yields with a maturity of two years or more, the difference in volatility decreases to only about 10 basis points.

M Additional Term Premium-Related Results

We present here additional results related to the term premium estimate obtained from the FS-ZLB model. In particular, we highlight how the FS-ZLB model induces structural changes in the market prices of risk during the ZLB regime, which in turn affect the dynamics of the latent factors.

Table 2: **Market Price of Risk Factor Loadings** Tabulated below are the loadings of the slope and curvature factors on the market price of risk in the FS-ZLB and FS-only models. λ_{it} represents the i th price of risk. Standard errors are in parentheses.

	FS-ZLB Model				FS-Only Model	
	Normal Regime		Lower Bound Regime		Slope	Curvature
	Slope	Curvature	Slope	Curvature		
λ_{1t}	-0.0012 (0.0362)	0.1421*** (0.0297)	-2.053*** (0.2045)	-2.002*** (0.1994)	-0.0394 (0.0315)	0.0496** (0.0241)
λ_{2t}	-0.0479 (0.0342)	-0.0277 (0.0263)	-0.0479 (0.0342)	-0.0277 (0.0263)	-0.0339 (0.0315)	0.132* (0.0686)
λ_{3t}	0.0091 (0.0341)	-0.0083 (0.0263)	0.0091 (0.0341)	-0.0083 (0.0263)	-0.0725*** (0.0134)	-0.0011 (0.0287)

Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

M.1 Structural Changes in the Market Price of Risk

In Equation (E.2), it is shown that the loadings of the factors on the market prices of risk depend on the difference between the mean reversion parameter $G_{s_t}^{\mathbb{P}}$ under the physical measure and the mean reversion parameter $G^{\mathbb{Q}}$ under the risk-neutral measure. Because the (1,2) and (1,3) entries of $G_{s_t}^{\mathbb{P}}$, denoted as $g_{s_t}^{\mathbb{P}}$, are equal to 0 during the lower bound regime, but the same entries are nonzero in $G^{\mathbb{Q}}$, the difference $G_{s_t}^{\mathbb{P}} - G^{\mathbb{Q}}$ should be larger during the lower bound regime than it is during the normal regime.

This is confirmed in Table 2, which compares the regime-wise loadings of the slope and curvature factors on the three prices of risk contained in $\lambda_{t,s_{t+1}}$; that is, it shows the last two columns of

$$\Lambda_{s_{t+1}} = \left(\Omega_{s_{t+1}}^{\frac{1}{2}} \right)^{\frac{1}{2}} \left(G_{s_{t+1}}^{\mathbb{P}} - G^{\mathbb{Q}} \right)$$

for $s_{t+1} = 1$, $s_{t+1} = 2$ and the FS-only model. The loadings of the two factors on the first market price of risk during the lower bound regime are more than 50 times larger than

their corresponding values in the normal regime. These loadings, whose absolute values exceed 1, are large even compared to those in the FS-only model, where no loading has magnitude greater than 1.

The change in $g_{s_t}^{\mathbb{P}}$ also reflects changes in short-term yield volatility. To see this, note that the first equation of the transition equation $f_t = G_{s_t}^{\mathbb{P}} f_{t-1} + \Omega_{s_t}^{\frac{1}{2}} v_t^{\mathbb{P}}$ can be written as

$$L_t = L_{t-1} + g_{s_t}^{\mathbb{P}'} f_{t-1}^{(2)} + \omega_{s_t} \cdot v_{1,t}^{\mathbb{P}}.$$

As such, as $g_{s_t}^{\mathbb{P}}$ increases, the correlation between the level factor and the (previous period's) slope and curvature factors rises, which can potentially ramp up the volatility of the level factor even if $\omega_{s_t}^2$ is small. By implication, the choice to put $g_{s_t}^{\mathbb{P}}$ equal to zero during the lower bound regime ensures that the level factor is minimally volatile when it is close to 0. Because the level factor loads most prominently on short-term yields, this has the added effect of compressing short-term yield volatility during the lower bound regime.

M.2 Bond Return Predictability

During the ZLB period, the evolution of term premium dynamics can be seen within the broader context of changes in the predictability of excess returns. Table 2 shows that the slope factor's influence on the market price of risk is insignificant during the normal regime. However, the curvature factor's influence on the first price of risk is significant in this regime, although its value is quite small, approximately 0.03. This indicates that yield spreads, as captured by the slope and curvature factors, can predict excess bond returns, which are calculated based on the market price of risk, during the normal regime. This supports the findings of Campbell and Shiller (1991), which highlight the ability of the yield spread to predict excess returns.

The yield spread continues to be predictable for excess returns in the lower bound regime, since Table 2 shows that the loadings of both the slope and curvature factors are highly significant when the lower bound is binding. However, the magnitude of the loadings is much larger during the lower bound regime than the normal regime. This seems to indicate that the yield spread has much greater predictive power concerning excess bond returns during ZLB episodes, which is exactly the result stated in Andreasen, Jørgensen, and Meldrum (2019).

There, it is demonstrated that the correlation between bond excess returns and the yield spread is much stronger when the lower bound is binding compared to when it is not. Andreasen, Jørgensen, and Meldrum (2019) also note that shadow rate models cannot replicate such changes in predictability, nor do they reflect the low level of short rate expectations following the GFC. Our model exhibits the ability to do both—the low level of short rate expectations after the GFC is implied by our high estimate of the term premium during this period—which may be considered an advantage of our model over the more popular shadow rate specification.

References

- Andreasen, M., K. Jørgensen, and A. Meldrum (2019). “Bond Risk Premiums at the Zero Lower Bound”.
- Bauer, Michael D. and Glenn D. Rudebusch (2020). “Interest Rates under Falling Stars”. In: *American Economic Review* 110.5, pp. 1316–54.
- Brand, C., G. Goy, and W. Lemke (2021). “Natural Rate Chimera and Bond Pricing Reality”.
- Campbell, John Y and Robert J Shiller (1987). “Cointegration and Tests of Present Value Models”. In: *Journal of Political Economy* 95.5, pp. 1062–1088.

- (1991). “Yield Spreads and Interest Rate Movements: A Bird’s Eye View”. In: *The Review of Economic Studies* 58.3, pp. 495–514.
- Cao, S. et al. (2020). *Fundamental Disagreement about Monetary Policy and the Term Structure of Interest Rates*. Tech. rep. FRB of New York.
- Cheridito, Patrick, Damir Filipovic, and Robert L. Kimmel (2007). “Market Price of Risk Specifications for Affine Models: Theory and Evidence”. In: *Journal of Financial Economics* 83.1, pp. 123–170.
- Chib, Siddhartha and Bakhodir Ergashev (2009). “Analysis of Multifactor Affine Yield Curve Models”. In: *Journal of the American Statistical Association* 104.488, pp. 1324–1337.
- Christensen, Jens H. E. (2013). *A Regime-Switching Model of the Yield Curve at the Zero Bound*. Tech. rep. FRB of San Francisco.
- Christensen, Jens H. E. and Glenn D. Rudebusch (2016). “Modeling Yields at the Zero Lower Bound: Are Shadow Rates the Solution?” In: *Advances in Econometrics* 35, pp. 75–125.
- Cinlar, Erhan (2011). *Probability and stochastics*. Vol. 261. Springer.
- Dai, Qiang and Kenneth J. Singleton (2000). “Specification Analysis of Affine Term Structure Models”. In: *The Journal of Finance* 55.5, pp. 1943–1978.
- Diebold, Francis X. and Canlin Li (2006). “Forecasting the Term Structure of Government Bond Yields”. In: *Journal of Econometrics* 130.2, pp. 337–364.
- Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007). “The US Treasury Yield Curve: 1961 to the Present”. In: *Journal of Monetary Economics* 54.8, pp. 2291–2304.
- Hall, Anthony D, Heather M Anderson, and Clive WJ Granger (1992). “A Cointegration Analysis of Treasury Bill Yields”. In: *The Review of Economics and Statistics*, pp. 116–126.

- Hamilton, James D and Jing Cynthia Wu (2012). “Identification and Estimation of Gaussian Affine Term Structure Models”. In: *Journal of Econometrics* 168.2, pp. 315–331.
- Hördahl, Peter and Oreste Tristani (2019). “Modelling Yields at the Lower Bound through Regime Shifts”. In: *Available at SSRN 3464245*.
- Johansen, Søren (1991). “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”. In: *Econometrica*, pp. 1551–1580.
- Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu (2011). “A New Perspective on Gaussian Dynamic Term Structure Models”. In: *The Review of Financial Studies* 24.3, pp. 926–970.
- Kopp, Emanuel and Peter D Williams (2018). *A Macroeconomic Approach to the Term Premium*. International Monetary Fund.
- Swanson, Eric T and John C Williams (2014). “Measuring the Effect of the Zero Lower Bound on Medium-and Longer-term Interest Rates”. In: *American Economic Review* 104.10, pp. 3154–3185.