# Decomposing Break-even Yields Into Inflation Expectations and Inflation Risk Premia

Serginio Sylvain

Current Draft: January, 2015
First Draft: January, 2015

# Table of Contents

## Contents

# 1 Theory

## 1.1 Break-even (BE) rate, expected inflation and inflation risk premium (IRP)

The derivation below is based on Christensen, Diebold and Rudebusch (2007). Let $\{P_t^N(\tau), P^R(\tau)\}$ denote the date $t$ prices of nominal and inflation-linked zero-coupon bonds (respectively) maturing at date $T$. we define $\tau \equiv T - t$ and $\{M_t^N, M_t^R\}$ denote the nominal and real stochastic discount factors.

$$
\begin{aligned}
P_t^N\{\tau\} &= E_t\left[\frac{M_{t+\tau}^N}{M_t^N} \times 1\right] = exp\left(-y_t^N\{\tau\} \cdot \tau\right) \\
P_t^R\{\tau\} &= E_t\left[\frac{M_{t+\tau}^R}{M_t^R} \times 1\right] = exp\left(-y_t^R\{\tau\} \cdot \tau\right) \\
y_t^N\{\tau\} &= -\frac{1}{\tau} ln\left(P_t^N\{\tau\}\right) \\
y_t^R\{\tau\} &= -\frac{1}{\tau} ln\left(P_t^R\{\tau\}\right) \\
y_t^N\{\tau\} - y_t^R\{\tau\} &= \frac{1}{\tau} ln\left(\frac{P_t^R\{\tau\}}{P_t^N\{\tau\}}\right) \\
\frac{dM_t^R}{M_t^R} &= -r_t^R dt - \Gamma_t \cdot dW_t^Q \\
\frac{dM_t^N}{M_t^N} &= -r_t^N dt - \Gamma_t \cdot dW_t^Q
\end{aligned}
$$

$$
\begin{aligned}
\text{Price level} &= \text{nominal consumption/real consumption} = \Pi_t \\
\Pi_t &= \frac{M_t^R}{M_t^N} \\
d(f\{x,y\}) &= f_x + f_y + \frac{1}{2}(f_{xx} + f_{yx} + f_{xy} + f_{yy}) \\
d(f\{x,y\}) &= f_x + f_y + \frac{1}{2}(f_{xx} + f_{yy}) + f_{xy} \\
d\Pi_t &= \frac{dM_t^R}{M_t^N} - \frac{M_t^R}{M_t^N}\left(\frac{dM_t^N}{M_t^N}\right) - \frac{dM_t^R}{M_t^N}\frac{dM_t^N}{M_t^N} + \frac{M_t^R}{M_t^N}\left(\frac{dM_t^N}{M_t^N}\right)^2 \\
\frac{d\Pi_t}{\Pi_t} &= \frac{dM_t^R}{M_t^R} - \frac{dM_t^N}{M_t^N} - \frac{dM_t^R}{M_t^N}\frac{dM_t^N}{M_t^N} + \left(\frac{dM_t^N}{M_t^N}\right)^2 \\
\frac{d\Pi_t}{\Pi_t} &= (r_t^N - r_t^R)dt \\
\Rightarrow dln(\Pi_t) &= (r_t^N - r_t^R)dt \\
\Rightarrow \Pi_t &= exp\left(\int_0^t (r_s^N - r_s^R)ds\right)
\end{aligned}
$$

$$
\begin{aligned}
P^N_t\{\tau\} &= E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\times\frac{\Pi_t}{\Pi_{t+\tau}}\right] \\
&= cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]+E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\right]E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right] \\
\frac{P^N_t\{\tau\}}{P^R_t\{\tau\}} &= \frac{cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]}{E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\right]}+E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right] \\
\frac{P^N_t\{\tau\}}{P^R_t\{\tau\}} &= E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]\left(\frac{cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]}{E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\right]E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]}+1\right) \\
-ln\left(\frac{P^R_t\{\tau\}}{P^N_t\{\tau\}}\right) &= ln\left(E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]\right)+ln\left(\frac{cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]}{E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\right]E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]}+1\right) \\
\Rightarrow \underbrace{y^N_t\{\tau\}-y^R_t\{\tau\}}_{\text{break-even rate}} &= \underbrace{\pi^e_t\{\tau\}}_{\text{expected inflation}} + \underbrace{\varphi_t\{\tau\}}_{\text{inflation risk premium (IRP)}} \\
\pi^e_t\{\tau\} &= -\frac{1}{\tau}ln\left(E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]\right) = -\frac{1}{\tau}ln\left(E_t\left[exp\left(-\int_t^{t+\tau}(r^N_s-r^R_s)ds\right)\right]\right) \\
\varphi_t\{\tau\} &= -\frac{1}{\tau}ln\left(\frac{cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]}{\underbrace{E_t\left[\frac{M^R_{t+\tau}}{M^R_t}\right]}_{(+)}\underbrace{E_t\left[\frac{\Pi_t}{\Pi_{t+\tau}}\right]}_{(+)}}+1\right)
\end{aligned}
$$

Notice that

> We observe positive inflation risk premiums [IRPs] if the real discount factor tends to be high (i.e., in a structural model, marginal utility is high) at the same time that price inflation is high (i.e., purchasing power is low) – Christensen, Lopez, Rudebusch (2010)

In other words, if future price increases are expected to occur in states of the world where consumers / investors are less wealthy (that is, in states of the world where consumers / investors are hungrier / derive more utility from marginal consumption) then $cov_t\left[\frac{M^R_{t+\tau}}{M^R_t},\frac{\Pi_t}{\Pi_{t+\tau}}\right]<0$ and the IRP is positive.

We have assumed that the ILB and the Nominal bond have the same liquidity. If that is not the case, the IRP will also contain the difference in liquidity premium for Nominal bonds versus ILBs. More specifically if ILBs are less liquid than Nominal bonds, the IRP ($\varphi_t\{\tau\}$) will be biased downwards. The market implied expected inflation ($\pi^e_t\{\tau\}$) may also be biased downwards if the real short rate is pushed up due to the illiquidity premium. To reduce this bias will will follow Christensen, Lopez and Rudebusch (2010) and use US TIPS data starting Jan 01, 2004. For US Nominal bonds we will use data starting Jan 01, 1995.

### 1.2 Nelson-Siegal arbitrage-free affine term structure (N-SAATS) model

We start with an affine term structure model as presented in Duffie and Kan (1996). Let's fix a probability space $(\Omega, \mathcal{F}, P)$ with filtration, $\{\mathcal{F}_t\}_{t\geq 0}$, generated by the four-dimensional standard Brownian Motion, $W_t$. The filtration satisfies the usual conditions. $P$ is the physical measure. Let $Q$ denote the risk neutral measure, $dW_t^Q = dW_t + \Gamma_t dt$. Here, $dW_t$ and $dW_t^Q$ are standard Brownian Motion increments under $P$ and $Q$, respectively[1]. Let $X_t$ denote the vector of state variables.

$$X_t = \begin{pmatrix} L_t^N \\ S_t \\ C_t \\ L_t^R \end{pmatrix}$$

$$dX_t = K_t^Q\left(\theta_t^Q - X_t\right)dt + \Sigma_t \begin{pmatrix} \sqrt{\nu_t^{(1)} + \beta_t^{(1)} \cdot X_t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\nu_t^{(4)} + \beta_t^{(4)} \cdot X_t} \end{pmatrix} dW_t^Q$$

To simplify the model, we will assume $\theta_t^Q = 0$, and $\nu_t^{(i)} = 1 \quad , \quad \beta_t^{(i)} = 0 \quad \forall i$. Following Christensen, Lopez, Rudebusch (2010), we assume that there are 4 state variables. Indeed, there is fair amount of evidence that level, slope and curvature factors work very well in explaining bond yields. In our model, $L^N$ and $L^R$ are the level factors from of Nominal and ILB respectively, $S_t$ and $C_t$ are slope and curvature factors. These last two are assumed to be common to both Nominal bonds and ILBs; this is again borrowed from Christensen, Lopez, Rudebusch (2010) who show that the slope factor for ILBs is highly correlated to that for the nominal bonds and similarly for the curvature factor).

To simplify the notation, let us suppress the $N$ and $R$ superscripts

$$P_t\{\tau\} = E_t^Q\left[exp\left(-\int_t^{t+\tau} r_s ds\right)\right]$$

$$r_t = \rho_0 + \rho_1 \cdot X_t \quad \text{[assuming short rates are affine in state variables]}$$

$$\Rightarrow P_t\{\tau\} = exp\left(B_t\{\tau\} \cdot X_t + G_t\{\tau\}\right)$$

Since $exp\left(-\int_0^t r_s ds\right) P_t\{\tau\}$ is Martingale under $Q$, it's drift is zero.

Thus, we allow for no arbitrage and

by using Ito's Lemman and setting $E_t^Q\left[\frac{dP_t\{\tau\}}{P_t\{\tau\}}\right] - r_t = 0$

$B_t\{\tau\}$ and $G_t\{\tau\}$ solve the ODEs in Section (**??**) of the Appendix

[1] $dW$ is a column vector, all other vectors are column vectors as well.

$$y_t\{\tau\} = -\frac{1}{\tau}ln\left(P_t\{\tau\}\right) = -\frac{1}{\tau}B_t\{\tau\}\cdot X_t - \frac{1}{\tau}G_t\{\tau\} \tag{1}$$

Solving for $B_t$ and $G_t$ typically requires imposing some parameter values and other restrictions with little motivation. Furthermore, finding the global solution in the estimation of the model parameters give the observed zero-coupon bond yields is often problematic (see Kim and Orphanides; 2005, Duffee; 2008). Instead, following Christensen, Lopez, Rudebusch (2010) we assume a dynamic Nelson-Seigel (1987) model and impose level, slope and curvature restrictions by replacing the Nelson-Seigel (1987) parameters with the level, slope and curvature state variables

$$\begin{aligned} y_t^N\{\tau\} &= L_t^N + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{G_t^N\{\tau\}}{\tau} \\ y_t^R\{\tau\} &= L_t^R + \alpha^R S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \alpha^R C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{G_t^R\{\tau\}}{\tau} \end{aligned}$$

We can motivate the Nelson-Seigel assumption from the fact that it is the workhorse model used in fitting yield curves. Furthermore, we need to use yields for zero-coupon Nominal bonds and ILBs of different maturities. Estimating these yields requires fitting a yield curve; this is again usually done with a Nelson-Seigel model. (1) then implies

$$\begin{aligned} B_t^N &= \begin{pmatrix} -(T-t) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ 0 \end{pmatrix} \\ B_t^R &= \begin{pmatrix} 0 \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ -(T-t) \end{pmatrix} \\ G_t^N\{\tau\} &= -\int_0^t \rho_0 + B_t' K_t^Q \theta_t^Q + \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t' B_t^N B_t^{N'}\Sigma_t\right)_{\{j,j\}}\nu_t^{N(j)} \\ G_t^R\{\tau\} &= -\int_0^t \rho_0 + B_t^{R'} K_t^Q \theta_t^Q + \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t' B_t^R B_t^{R'}\Sigma_t\right)_{\{j,j\}}\nu_t^{R(j)} \end{aligned}$$

$$P_T\{0\} = 1 \quad \Rightarrow \quad B_T\{0\} = G_T\{0\} = 0$$

Hence, given the Nelson-Seigel (N-S) restriction, there exists a unique class of Affine Term Structure models[2] which satisfies the ODEs for$B_t$ and $G_t$ in Section (**??**) of the Appendix:

[2]As shown in Section 4.4 of the Appendix, we can set $\theta^Q = 0$ without loss of generality.

$$
\begin{aligned}
X_t &= \begin{pmatrix} L_t^N \\ S_t \\ C_t \\ L_t^R \end{pmatrix} \\
dX_t &= -\underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & -\lambda & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{K^Q} X_t dt + \Sigma dW_t^Q \\
r_t^N &= \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{\rho^N} \cdot X_t = L_t^N + S_t \\
r_t^R &= \underbrace{\begin{pmatrix} 0 \\ \alpha^R \\ 0 \\ 1 \end{pmatrix}}_{\rho^R} \cdot X_t = L_t^R + \alpha^R S_t \\
\begin{pmatrix} \frac{d}{dt}B_{1,t}^N \\ \frac{d}{dt}B_{2,t}^N \\ \frac{d}{dt}B_{3,t}^N \\ \frac{d}{dt}B_{4,t}^N \end{pmatrix} &= \begin{pmatrix} 1 \\ e^{-\lambda(T-t)} \\ \lambda e^{-\lambda(T-t)}(T-t) \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 + B_{2,t}^N \lambda \\ B_{3,t}^N \lambda - B_{2,t}^N \lambda \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{\rho^N} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{K^{Q\prime}} \begin{pmatrix} B_{1,t}^N \\ B_{2,t}^N \\ B_{3,t}^N \\ B_{4,t}^N \end{pmatrix} \\
\frac{d}{dt} G_t^N &= -\frac{1}{2} \sum_{j=1}^{4} \left( \Sigma B_t^N B_t^{N\prime} \Sigma \right)_{\{j,j\}} \\
B_T^N &= G_T^N = 0 \\
\frac{d}{dt} B_t^R &= \begin{pmatrix} 0 \\ \alpha^R e^{-\lambda(T-t)} \\ \alpha^R \lambda e^{-\lambda(T-t)}(T-t) \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ B_{2,t}^R \lambda \\ B_{3,t}^R \lambda - B_{2,t}^R \lambda \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ \alpha^R \\ 0 \\ 1 \end{pmatrix}}_{\rho^R} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{K^{Q\prime}} \begin{pmatrix} B_{1,t}^R \\ B_{2,t}^R \\ B_{3,t}^R \\ B_{4,t}^R \end{pmatrix} \\
\frac{d}{dt} G_t^R &= -\frac{1}{2} \sum_{j=1}^{4} \left( \Sigma' B_t^R B_t^{R\prime} \Sigma \right)_{\{j,j\}} \\
B_T^R &= G_T^R = 0
\end{aligned}
$$

Notice that this class of models does away with the square root process for the state variables so technically $X_t$ can be negative. This is a drawback of the models nonetheless, as shown in Christensen, Lopez and Rudebusch (2010) the estimation is unlikely to yield negative values for $X_t$. Furthermore, we hope to apply these models European countries as well; and we have observed negative nominal yields for some European government bonds.

To bring the model to data we need to switch to the $P$ measure

$$
\begin{aligned}
dX_t &= K^P\left(\theta^P - X_t\right)dt + \Sigma dW_t \\
y_t^N &= -\frac{1}{T-t}B_t^N \cdot X_t - \frac{1}{T-t}G_t^N \\
y_t^R &= -\frac{1}{T-t}B_t^R \cdot X_t - \frac{1}{T-t}G_t^R
\end{aligned}
$$

Notice that $K^P$ and $\theta^P$ are unrestricted (see Section 4.4 of the Appendix for the derivation). Thus, the model allows for plenty of flexibility. And, the state variables will be (unconditionally) correlated even if if we impose that $\Sigma$ is diagonal.

In the model key equations below, Variations in $dX_t$ and $X_t$ allows to identify $K^P$. Then, the unconditional mean of $X_t$ provide the four equations needed to identify . Variations in $y_t^N$,$X_t$ and $\tau$ allow us to identify $\lambda$and then variations in $y_t^R$, $X_t$ and$\tau$ allow us to identify $\alpha^R$. Lastly, variations in $y_t^N$, $y_t^R$ and $\tau$ allow us to identify $N(N+1)/2$ elements of $\Sigma$ as shown in Christensen, Diebold and Rudebusch (2007). For simplicity we assume that $\Sigma$ is diagonal. The key equations for the model become

$$
\begin{pmatrix} dL_t^N \\ dS_t \\ dC_t \\ dL_t^R \end{pmatrix} = \begin{pmatrix} K_{11}^P & K_{12}^P & K_{13}^P & K_{14}^P \\ K_{21}^P & K_{22}^P & K_{23}^P & K_{24}^P \\ K_{31}^P & K_{32}^P & K_{33}^P & K_{34}^P \\ K_{41}^P & K_{42}^P & K_{43}^P & K_{44}^P \end{pmatrix} \left( \begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \\ \theta_4^P \end{pmatrix} - \begin{pmatrix} L_t^N \\ S_t \\ C_t \\ L_t^R \end{pmatrix} \right) dt + \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix} dW_t \qquad (2
$$

$$
\begin{aligned}
y_t^N &= -\frac{1}{T-t}\begin{pmatrix} -(T-t) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ 0 \end{pmatrix}' \begin{pmatrix} L_t^N \\ S_t \\ C_t \\ L_t^R \end{pmatrix} - \frac{\sigma_{11}^2}{2(T-t)}\left(\int_t^T (T-s)^2 ds\right) \\
&\quad -\frac{\sigma_{22}^2}{2(T-t)}\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda}\right)^2 ds\right) - \frac{\sigma_{33}^2}{2(T-t)}\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda} - e^{-\lambda(T-s)}(T-s)\right)^2 ds\right) \qquad (3
\end{aligned}
$$

$$
\begin{aligned}
y_t^R &= -\frac{1}{T-t}\begin{pmatrix} 0 \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ -(T-t) \end{pmatrix}' \begin{pmatrix} L_t^N \\ S_t \\ C_t \\ L_t^R \end{pmatrix} - \frac{\sigma_{22}^2}{2(T-t)}(\alpha^R)^2\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda}\right)^2 ds \right. \\
&\quad -\frac{\sigma_{33}^2}{2(T-t)}(\alpha^R)^2\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda} - e^{-\lambda(T-s)}(T-s)\right)^2 ds\right) - \frac{\sigma_{44}^2}{2(T-t)}\left(\int_t^T (T-s)^2 ds\right) \qquad (4
\end{aligned}
$$

## 1.3 Implications

In Section 4.7 of the Appendix we derive

$$
\begin{aligned}
ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right) &= \int_t^{t+\tau}(\rho^N-\rho^R)' X_s ds \\
\frac{\partial}{\partial\tau} ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right) &= (\rho^N-\rho^R)' X_{t+\tau} \\
Z_{t,t+\tau} &\equiv \begin{pmatrix} X_{t+\tau} \\ ln\left(\Pi_{t+\tau}/\Pi_t\right)\end{pmatrix} \qquad \text{the augmented state vector} \\
m_{t,t+\tau} &= E_t\left[Z_{t,t+\tau}\right] \\
\frac{\partial}{\partial\tau} m_{t,t+\tau} &= K+\theta m_{t,t+\tau} \\
m_{t,t} &= Z_{t,t} = \begin{pmatrix} X_t \\ 0 \end{pmatrix} \\
v\left\{\tau\right\} &= Var_t\left[Z_{t,t+\tau}\right] \\
\frac{\partial}{\partial\tau} vec\left(v\left\{\tau\right\}\right) &= \left(I_{n\times n}\otimes\theta\right) vec\left(v\left\{\tau\right\}\right) + \left(\theta\otimes I_{n\times n}\right) vec\left(v\left\{\tau\right\}\right) + \left(I_{n\times n}\otimes\bar{\Sigma}\right) vec\left(\bar{\Sigma}'\right) \\
vec\left(v\left\{0\right\}\right) &= 0_{n^2\times 1} \quad \text{where } n \text{ is the size of the augmented state vector}
\end{aligned}
$$

We then have

$$
\begin{aligned}
\pi_t^e\left\{\tau\right\} &= -\frac{1}{\tau} ln\left(E_t\left[exp\left(-\int_t^{t+\tau}(r_s^N-r_s^R)ds\right)\right]\right) = -\frac{1}{\tau} ln\left(E_t\left[exp\left(-(\rho^N-\rho^R)'\int_t^{t+\tau} X_s ds\right)\right]\right) \\
\Rightarrow \pi_t^e\left\{\tau\right\} &= -\frac{1}{\tau}\left(-m_{t,t+\tau}^{(n)}+\frac{1}{2}v^{(n,n)}\left\{\tau\right\}\right) \\
\pi_t^e\left\{0\right\} &= (\rho^N-\rho^R)' X_t
\end{aligned}
$$

### 1.3.1 Probability of deflation

Probability that at date $t$, inflation over horizon $\tau$ is negative

$$
\begin{aligned}
Prob_t\left(ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right)\leq 0\right) &= Prob_t\left(\int_t^{t+\tau}(\rho^N-\rho^R)' X_s ds \leq 0\right) \\
Prob_t\left(ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right)\leq 0\right) &= Prob_t\left(\frac{\int_t^{t+\tau}(\rho^N-\rho^R)' X_s ds - m_{t,t+\tau}^{(n)}}{\sqrt{v^{(n,n)}\left\{\tau\right\}}} \leq \frac{-m_{t,t+\tau}^{(n)}}{\sqrt{v^{(n,n)}\left\{\tau\right\}}}\right)
\end{aligned}
$$

$$Prob_t\left(ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right)\leq 0\right) = CDF\left\{\frac{-m^{(n)}_{t,t+\tau}}{\sqrt{v^{(n,n)}\left\{\tau\right\}}}\right\}$$

We should not that this is also the probability that at date $t$, average inflation over horizon $\tau$ is negative

$$Prob_t\left(\frac{1}{\tau}ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right)\leq 0\right) = Prob_t\left(ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right)\leq 0\right) = CDF\left\{\frac{-m^{(n)}_{t,t+\tau}}{\sqrt{v^{(n,n)}\left\{\tau\right\}}}\right\}$$

## 1.4 Extending the number of states

### 1.4.1 Additional hidden states

The model with four state variables struggles to fit the short maturity bonds. To make it a bit more flexible we would need to introduce one more maturity parameter $\lambda_2$. This would require introducing two new stat variables $S_{2,t}$and $C_{2,t}$.

$$y^N_t\{\tau\} = L^N_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + S_{2,t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + C_{2,t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\right) - \frac{G^N_t\left\{\tau\right\}}{\tau}$$

$$y^R_t\{\tau\} = L^R_t + \alpha^R S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \alpha^R S_{2,t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau}\right) + \alpha^R C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \alpha^R C_{2,t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\right) -$$

### 1.4.2 Introducing observed macro-economic and/or technical state variables

Suppose we use 4 unobserved state variables and some observed variables, $\tilde{X}_t$.

$$X_t = \begin{pmatrix} L^N_t \\ S_t \\ C_t \\ L^R_t \\ \tilde{X}_t \end{pmatrix}$$

$$y^N_t\{\tau\} = L^N_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{G^N_t\left\{\tau\right\}}{\tau} + \Lambda_N\tilde{X}_t$$

$$y^R_t\{\tau\} = L^R_t + \alpha^R S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \alpha^R C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{G^R_t\left\{\tau\right\}}{\tau} + \Lambda_R\tilde{X}_t$$

We then have

$$B^R_{5,t}\left\{\tau\right\} = -\tau\Lambda_R$$

$$B^N_{5,t}\left\{\tau\right\} = -\tau\Lambda_N$$

$$K^Q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\rho^N = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \Lambda_N \end{pmatrix}$$

$$\rho^R = \begin{pmatrix} 0 \\ \alpha^R \\ 0 \\ 1 \\ \Lambda_R \end{pmatrix}$$

We would need only a slight modification the measurement equation of the Kalman Filter.

$$\begin{aligned} Y_t = \begin{pmatrix} y_t \\ \tilde{X}_t \end{pmatrix} &= A_0 + A_1 X_t + \epsilon_t \\ \epsilon_t^N &\sim N(0, \Phi) \\ A_1 &= \begin{pmatrix} A_1^N \\ A_1^R \\ [0,0,0,0,1] \end{pmatrix}, \qquad A_0 = \begin{pmatrix} A_0^N \\ A_0^R \\ 0 \end{pmatrix} \\ \Phi &= \begin{pmatrix} \phi_N^2 & 0 & 0 \\ 0 & \phi_R^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

The subsequent steps of the Kalman Filter estimation algorithm remain unchanged.

# 2 Estimation Method

From the key equations (2)-(4) we need to estimate 26+ parameters (depending on the number of bond maturities included). $X_t$ is not observed . We can use a filtering method to estimate both $X_t$ and the parameters. This is the approached used in Christensen, Diebold and Rudebusch (2007) and Christensen, Lopez and Rudebusch (2010).

## 2.1 Kalman filter

Here the state variables are assumed to be hidden.

When using daily data we may have asynchronous data releases[3]. That is more so the case if we include macro-economic variables as additional state variables. A Kalman filter approach would be appropriate in that case. Recall that since the all the shocks are Gaussian, Kalman-filtering is an efficient and consistent estimator.

From Section 4.5 of the Appendix, the discretized state equation is

$$X_i = U_0 + U_1 X_{t_i} + \eta_i$$

where

[3] Indeed since each yield at each maturity has its own iid error, we can simply drop the rows/columns corresponding to the missing data.

$$\Delta = t_i - t_{i-1} = \frac{1}{252} \quad \text{for daily data}$$

$$\begin{aligned} U_0 &= \left(I - exp\left(-K^p \Delta t_i\right)\right)\theta^P \\ U_1 &= exp\left(-K^p \Delta t_i\right) \end{aligned}$$

$$\eta_i \sim N(0, Q)$$

$$Q = \int_0^\Delta exp\left(-K^p s\right)\Sigma\Sigma' exp\left(-K^p s\right)' ds$$

The measurement equation is obtained from equations (3)-(4) where $A_0$and $A_1$are appropriately defined and $\Phi$ is diagonal. We will use several maturities of up to $M$ years ($T \in \{T_1, ..., T_M\}$ with $T_n = \frac{n}{\Delta}$; hence $\tau \in \{\tau_1, ..., \tau_M\}$ with $\tau_n = T_n - t$).

$$y_t = A_0 + A_1 X_t + \epsilon_t$$

$$\epsilon_t^N \sim N(0, \Phi)$$

$$A_1 = \begin{pmatrix} A_1^N \\ A_1^R \end{pmatrix}, \qquad A_0 = \begin{pmatrix} A_0^N \\ A_0^R \end{pmatrix}$$

$$A_1^N\{\tau\} = -\frac{1}{\tau}\begin{pmatrix} -\tau \\ -\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right) \\ -\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\ 0 \end{pmatrix}$$

$$\begin{aligned} A_0^N\{\tau\} = &-\frac{\sigma_{11}^2}{2\tau}\left(\frac{\tau^3}{3}\right) \\ &-\frac{\sigma_{22}^2}{2\tau}\left(-\frac{-2\lambda\tau + e^{-2\lambda\tau} - 4e^{-\lambda\tau} + 3}{2\lambda^3}\right) \\ &-\frac{\sigma_{33}^2}{2\tau}\left(\frac{e^{-2\lambda\tau}\left(8e^{\lambda\tau}(\lambda\tau+2) - 2\lambda\tau(\lambda\tau+3) + e^{2\lambda\tau}(4\lambda\tau - 11) - 5\right)}{4\lambda^3}\right) \end{aligned}$$

$$A_1^R\{\tau\} = -\frac{1}{\tau}\begin{pmatrix} 0 \\ -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right) \\ -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\ -\tau \end{pmatrix}$$

$$\begin{aligned} A_0^R\{\tau\} = &-\frac{\sigma_{22}^2}{2\tau}(\alpha^R)^2\left(-\frac{-2\lambda\tau + e^{-2\lambda\tau} - 4e^{-\lambda\tau} + 3}{2\lambda^3}\right) \\ &-\frac{\sigma_{33}^2}{2\tau}(\alpha^R)^2\left(\frac{e^{-2\lambda\tau}\left(8e^{\lambda\tau}(\lambda\tau+2) - 2\lambda\tau(\lambda\tau+3) + e^{2\lambda\tau}(4\lambda\tau - 11) - 5\right)}{4\lambda^3}\right) \end{aligned}$$

$$-\frac{\sigma_{44}^2}{2\tau}\left(\frac{\tau^3}{3}\right)$$

To avoid over-fitting the model we assume that $\Phi$is diagonal.

### 2.1.1 Algorithm

1. Initialization: we will initialize using the unconditional mean and variance[4]

$$\begin{aligned}\hat{V}_{0|0} &= \int_0^{\infty} exp\left(-K^p s\right)\Sigma\Sigma' exp\left(-K^p s\right)' ds \\ \hat{X}_{0|0} &= \theta^P\end{aligned}$$

   For stationarity ($V_0^{(i,j)} < \infty, \forall\{i,j\}$) we require that the real component of the eigenvalues of $K^p$ are positive.

2. Predicted states and error variance

$$\begin{aligned}\hat{X}_{k|k-1} &= U_0 + U_1\hat{X}_{k-1|k-1} \\ \hat{V}_{k|k-1} &= U_1\hat{V}_{k-1|k-1}U_1' + Q\end{aligned}$$

3. Update

$$\begin{aligned}\varepsilon_k &= y_k - A_0 - A_1\hat{X}_{k|k-1} && \text{innovation} \\ S_k &= \Phi + A_1\hat{V}_{k|k-1}A_1' && \text{innovation variance} \\ G_k &= \hat{V}_{k|k-1}A_1'(S_k)^{-1} && \text{Kalman gains (see Appendix for derivation)} \\ \hat{X}_{k|k} &= \hat{X}_{k|k-1} + G_k\varepsilon_k \\ \hat{V}_{k|k} &= \hat{V}_{k|k-1} - G_kA_1\hat{V}_{k|k-1}\end{aligned}$$

## 2.2 E-M Algorithm

To find the optimal parameters we use the EM algorithm. We will use two versions of the EM algorithm, the first is implemented with MLE and the second uses a Bayesian approach.

[4] In our implementation we use, $V_0 = \int_0^{1000} exp\left(K^p s\right)\Sigma\Sigma' exp\left(K^p s\right)' ds$

Notice that we can stack $Y_t$ and $X_t$

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} A_0 \\ U_0 \end{pmatrix} + \begin{pmatrix} A_1 & 0 \\ 0 & U_1 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left( \left[ \begin{pmatrix} Y_t \\ X_t \end{pmatrix} - \begin{pmatrix} A_0 \\ U_0 \end{pmatrix} - \begin{pmatrix} A_1 & 0 \\ 0 & U_1 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} \right], \begin{bmatrix} \Phi & 0 \\ 0 & Q \end{bmatrix} \right)$$

Since we assume $\epsilon_t$ and $\eta_t$ are independent, the joint likelihood is

$$\prod_{t=1}^{T} p\left(Y_t, X_t | X_{t-1}, \Omega\right) = \left( \prod_{t=1}^{T} p\left(Y_t | X_t, \Omega\right) \right) \left( \prod_{t=1}^{T} p\left(X_t | X_{t-1}, \Omega\right) \right)$$

where $\Omega$ is the set of parameters

#### 2.2.1 MLE

Below is a version of the EM algorithm[5].

First, note that

$$\begin{aligned}
\mathcal{L}\{\Omega\} &= ln\left( \prod_{t=1}^{T} p\left(Y_t | \Omega\right) \right) \\
\mathcal{L}\{\Omega\} &= ln\left( \prod_{t=1}^{T} \left[ \int_{X_t} p\left(Y_t, X_t | X_{t-1}, \Omega\right) dX_t \right] \right) \\
\mathcal{L}\{\Omega\} &= ln\left( \prod_{t=1}^{T} E_{X_t}\left[ p\left(Y_t, X_t | X_{t-1}, \Omega\right) \right] \right) \\
\mathcal{L}\{\Omega\} &= \sum_{t=1}^{T} ln\left( E_{X_t}\left[ p\left(Y_t, X_t | X_{t-1}, \Omega\right) \right] \right) \\
& \quad \text{From Jensens' Inequality} \\
\mathcal{L}\{\Omega\} &\geq \sum_{t=1}^{T} E_{X_t}\left[ ln\left( p\left(Y_t | X_t, \Omega\right) p\left(X_t | X_{t-1}, \Omega\right) \right) \right] \\
& \quad \text{By the the Law of Iterate Expectations} \\
\mathcal{L}\{\Omega\} &\geq E\left[ \sum_{t=1}^{T} ln\left( p\left(Y_t | X_t, \Omega\right) p\left(X_t | X_{t-1}, \Omega\right) \right) \middle| X^{(T)} \right] \\
\mathcal{L}\{\Omega\} &\geq E\left[ ln\left( \left( \prod_{t=1}^{T} p\left(Y_t | X_t, \Omega\right) \right) \left( \prod_{t=1}^{T} p\left(X_t | X_{t-1}, \Omega\right) \right) \right) \middle| X^{(T)} \right]
\end{aligned}$$

To avoid over-fitting we set $X_0 = \theta^p$

1. Start with a guess for the set of parameters,$\Omega$

2. Run the Kalman filter

3. Run the Kalman smoother

[5] See slide 152 of http://www.ssc.upenn.edu/~fdiebold/Teaching706/TimeSeriesSlides.pdf

4. Compute the expected log-likelihood. This the lower bound to $\mathcal{L}\left\{\Omega\right\}$and is the objective function we will maximize

$$\mathscr{L}\{\Omega\} \quad = \quad E\left[ln\left(\left(\prod_{t=1}^{T}p\left(Y_t|X_t,\Omega\right)\right)\left(\prod_{t=1}^{T}p\left(X_t|X_{t-1},\Omega\right)\right)\right)\Big|X^{(T)}\right]$$

We (may) impose the non-linear constraints that the real component of each eigenvalues of $K^p$ positive (so $K^p$ is positive definite).

5. Solve for $\hat{\Omega}$

$$\hat{\Omega} \quad = \quad arg\max_{\Omega}\mathscr{L}\{\Omega\}$$

6. repeat steps 2-5 until change in $\hat{\Omega}$ is below a pre-specified tolerance level

### 2.2.2 Bayesian estimation

Below is a version of the Variational Bayesian EM algorithm[6].

To avoid over-fitting we set $X_0 = \theta^p$

1. Start with a guess for the set of parameters,$\Omega$

2. Run the Kalman filter

3. Run the Kalman smoother

4. Choose priors for $\Omega$

5. Compute the expected log-likelihood

$$\mathscr{L}\{\Omega\} \quad = \quad E\left[ln\left(\left(\prod_{t=1}^{T}p\left(Y_t|X_t,\Omega\right)\right)\left(\prod_{t=1}^{T}p\left(X_t|X_{t-1},\Omega\right)\right)\right)\Big|X^{(T)}\right]$$

6. Find posterior for $\Omega$ via MCMC or other sampling methods

7. repeat steps 2-6 until change in the mean of the parameters is below a pre-specified tolerance level

## 2.3 Forecasts

Given

$$\begin{aligned}
y_t &= A_0 + A_1 X_t + \epsilon_t \\
X_t &= U_0 + U_1 X_{t-1} + \eta_t
\end{aligned}$$

Conditional on any date $t$ we can forecast $y_t\left\{h\right\} = E_t\left[y_{t+h}\right]$ for any $h > 0$

$$\begin{aligned}
X_{t+h} &= U_0 + U_1 X_{t+h-1} + \eta_{t+h} \\
X_{t+h} &= U_0 + U_1\left(U_0 + U_1 X_{t+h-2} + \eta_{t+h-1}\right) + \eta_{t+h} \\
X_{t+h} &= U_0 + U_1 U_0 + U_1^2 X_{t+h-2} + U_1\eta_{t+h-1} + \eta_{t+h} \\
X_{t+h} &= \sum_{j=0}^{h-1} U_1^j U_0 + U_1^h X_t + \sum_{j=0}^{h-1} U_1^j \eta_{t+h-j}
\end{aligned}$$

[6]See http://www.cse.buffalo.edu/faculty/mbeal/thesis/beal03_2.pdf

$$y_{t+h} = A_0 + A_1\left(\sum_{j=0}^{h-1} U_1^j U_0 + U_1^h X_t + \sum_{j=0}^{h-1} U_1^j \eta_{t+h-j}\right) + \epsilon_{t+h}$$

The forecast error is

$$y_{t+h} - y_t\{h\} = A_1 \sum_{j=0}^{h-1} U_1^j \eta_{t+h-j}$$

The forecast error variance

$$\varsigma_t = E_t\left[\left(y_{t+h} - y_t\{h\}\right)\left(y_{t+h} - y_t\{h\}\right)'\right] = E_t\left[\left(A_1 \sum_{j=0}^{h-1} U_1^j \eta_{t+h-j}\right)\left(\left(\sum_{j=0}^{h-1} \eta_{t+h-j}'\left(U_1^j\right)'\right) A_1'\right)\right]$$

$$\varsigma_t\{h\} = A_1\left(\sum_{j=0}^{h-1} U_1^j Q\left(U_1^j\right)'\right) A_1'$$

$$\varsigma_t\{h\} = \varsigma_t\{h-1\} + A_1 U_1^{h-1} Q\left(U_1^{h-1}\right)' A_1'$$

$$\varsigma_t\{1\} = A_1 Q A_1'$$

### 2.4 Data

Ideally we would like to use raw zero-coupon yields. These can be obtained from the Fama and Bliss (1987)

> who construct yields not from an estimated discount curve, but rather from estimated forward rates at the observed maturities. Their method sequentially constructs the forward rates necessary to price successively longer- maturity bonds. Those forward rates are often called "unsmoothed Fama-Bliss" forward rates, and they are transformed to unsmoothed Fama-Bliss yields by appropriate averaging[7] – "Facts, Factors, and Questions" (Diebold and Rudebusch; 2013)

Unfortunately, the Fama and Bliss (1987) data is only available monthly and for maturities of up to five years. Furthermore it does not contain the yields for ILBs. Consequently, we will instead use the smoothed zero-coupon yields provided Gurkaynak, Sack and Wright (2006) for Nominal Bonds and Gurkaynak, Sack and Wright (2008) for ILBs[8]. This does introduce some error because in their construction Gurkaynak, Sack and Wright (2006, 2008) exclude the on-the-run bonds. Hence the Nominal Bond yields and ILB yields are biased upwards. However, since we are interested in studying the Break-Even rate this is not problematic so long as the on-the-run premium is roughly the same in Nominal Bonds and ILBs. If that is not the case, part of the IRP will reflect this difference. What is more worrisome however is that Gurkaynak, Sack and Wright (2006, 2008) use an Nelson-Seigal approach to smooth the yield curve. Thus if the yield curve were to remain constant (over time), our estimation procedure would simply recover the parameters of Gurkaynak, Sack and Wright (2006, 2008). To mitigate this issue we will only use maturies for which their were/are outstanding Treasury securities that were/are used by Gurkaynak, Sack and Wright (2006, 2008) to fit the yield curve at any point in time. Hence, these smoothed yields are close to the observed yields. Lastly to test whether our estimation method is robust we can fix the Gurkaynak, Sack and Wright (2006, 2008) parameters (hold the yield curve

[7]The zero-coupon yield is the equal-weighted mean of the forward rates. Hence, it prices all bonds of lesser or equal maturities.

[8]This data is available publicly, is of daily frequency and is updated regularly. See http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html, http://www.federalreserve.gov/pubs/feds/2008/200805/200805abs.ht http://www.federalreserve.gov/econresdata/researchdata/feds200628.xls, and http://www.federalreserve.gov/econresdata/researchdata/feds20

constant overtime) and verify that our procedure does recover the Gurkaynak, Sack and Wright (2006, 2008) parameters.

### 2.5 Robustness check with GMM

Having estimated the model's parameters, we will use the time series of Nominal bond and ILB returns and verify whether the model implied SDF's price these bonds. In particular, we will use the moment conditions from the Euler Equations

$$
\begin{aligned}
E_t\left[\frac{M^N_{t+k}}{M^N_t} R^N_{t,t+k}\{\tau\}\right] &= 1 \\
E_t\left[\frac{M^R_{t+k}}{M^R_t} R^R_{t,t+k}\{\tau\}\right] &= 1
\end{aligned}
$$

and use the J-test to verify if the model can price these bonds. Since we have estimated $l \sim 30$ parameters we need $m > l$ moment conditions for the J-Test. Thus we need to use many different maturities for Nominal bonds and ILBs

$$
J = \begin{pmatrix} \frac{1}{K}\sum_t^K \left(\frac{M^N_{t+k}}{M^N_t} R^N_{t,t+k}\right) \\ \vdots \\ \frac{1}{K}\sum_t^K \left(\frac{M^R_{t+k}}{M^R_t} R^R_{t,t+k}\right) \end{pmatrix}' \begin{pmatrix} \frac{1}{K}\sum_t^K \left(\frac{M^N_{t+k}}{M^N_t} R^N_{t,t+k}\right) \\ \vdots \\ \frac{1}{K}\sum_t^K \left(\frac{M^R_{t+k}}{M^R_t} R^R_{t,t+k}\right) \end{pmatrix} \sim \chi^2_{m-l}
$$

We can also include total returns for TIPS ETF, Nominal Bond ETFs, realized total returns for active Treasuries of maturities $1-3Y$, $5-7Y$, $7-10Y$, $10Y+$, realized total returns for active TIPS of maturities $1-3Y$, $5-7Y$, $7-10Y$, $10Y+$, for horizons $k \in \{1D, 1W, 1M\}$ (this should yield a total of more than 30 test assets).

### 2.6 Signal construction

### 2.7 Extensions

Extension to other countries/regions

### 2.8 Alternative approach

If we instead assume $X_t$ is observed (estimated from PCA), we can first stack $X_t$ and $y_t$

$$
\begin{aligned}
Y_t &= \begin{pmatrix} X_t \\ y_t \end{pmatrix} \\
Y_t &= B_0 + B_1 Y_{t-1} + \mu_t \\
\mu_t &\sim N\left(0, \begin{pmatrix} V & 0 \\ 0 & \Phi \end{pmatrix}\right)
\end{aligned}
$$

we could then use MLE. The standard errors may need to be corrected for errors coming from the PCA. This can be done using a Bootstrap approach but would be very time consuming.

# 3

## 3.1 Missing data

If we use macro-economic variables as additional state variables there may be some missing, delayed or intermittent data. The Kalman Filter can handle these cases. In particular, when $\tilde{X}_t$ is missing

$$\Phi = \begin{pmatrix} \phi_N^2 & 0 & 0 \\ 0 & \phi_R^2 & 0 \\ 0 & 0 & \infty \end{pmatrix}$$

and the corresponding elements of the update step of the Kalman filter become

$$\begin{aligned}
S_k\{5\} &= \infty \qquad \text{innovation variance} \\
G_k\{5\} &= 0 \qquad \text{Kalman gains (see Appendix for derivation)} \\
\hat{X}_{k|k}\{5\} &= \hat{X}_{k|k-1}\{5\} \\
\hat{V}_{k|k}\{5\} &= \hat{V}_{k|k-1}\{5\}
\end{aligned}$$

# 4 Appendix

## 4.1 Deriving ODEs for $B_t$ and $G_t$

$$\begin{aligned}
exp\left(-\int_0^t r_s ds\right) P_t\{\tau\} &= exp\left(B_t^{'}\{\tau\} X_t + G_t\{\tau\} - \int_0^t r_s ds\right) \\
E_t^Q\left[\frac{dP_t\{\tau\}}{P_t\{\tau\}}\right] - r_t &= -\rho_0 - \rho_1 \cdot X_t + X_t \cdot \frac{\partial}{\partial t} B_t\{\tau\} dt + \frac{\partial}{\partial t} G_t\{\tau\} dt + B_t^{'}\{\tau\} E_t^Q[dX_t] \\
&\quad + \frac{1}{2}\left(B_t\{\tau\}^{'}(dX_t dX_t^{'}) B_t\{\tau\}\right) \\
E_t^Q\left[\frac{dP_t\{\tau\}}{P_t\{\tau\}}\right] - r_t &= 0 \\
\Rightarrow 0 &= -\rho_0 - \rho_1 \cdot X_t + X_t \cdot \frac{d}{dt} B_t\{\tau\} dt + \frac{d}{dt} G_t\{\tau\} dt + B_t\{\tau\}^{'} K_t^Q\left(\theta_t^Q - X_t\right) dt \\
&\quad + \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t^{'} B_t\{\tau\} B_t\{\tau\}^{'} \Sigma_t\right)_{\{j,j\}}\left(\nu_t^{(j)} + \beta_t^{(j)} \cdot X_t\right) dt \\
\Rightarrow \frac{d}{dt} B_t\{\tau\} &= \rho_1 + K_t^{Q'} B_t\{\tau\} - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t^{'} B_t\{\tau\} B_t\{\tau\}^{'} \Sigma_t\right)_{\{j,j\}}\left(\beta_t^{(j)}\right) && (5) \\
\Rightarrow \frac{d}{dt} G_t\{\tau\} &= \rho_0 - B_t\{\tau\}^{'} K_t^Q \theta_t^Q - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t^{'} B_t\{\tau\} B_t\{\tau\}^{'} \Sigma_t\right)_{\{j,j\}} \nu_t^{(j)} && (6) \\
P_t\{0\} = 1 &\Rightarrow B_t\{0\} = G_t\{0\} = 0 && (7)
\end{aligned}$$

## 4.2 Original Nelson-Seigel (1987) model

Recall the original Nelson-Seigel (1987) model

$$y_t\{\tau\} = \underbrace{a_0}_{\text{long term}} + \underbrace{a_1\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)}_{\text{short term}} + \underbrace{a_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)}_{\text{medium term}}$$

Assuming $\lambda = 0.5$

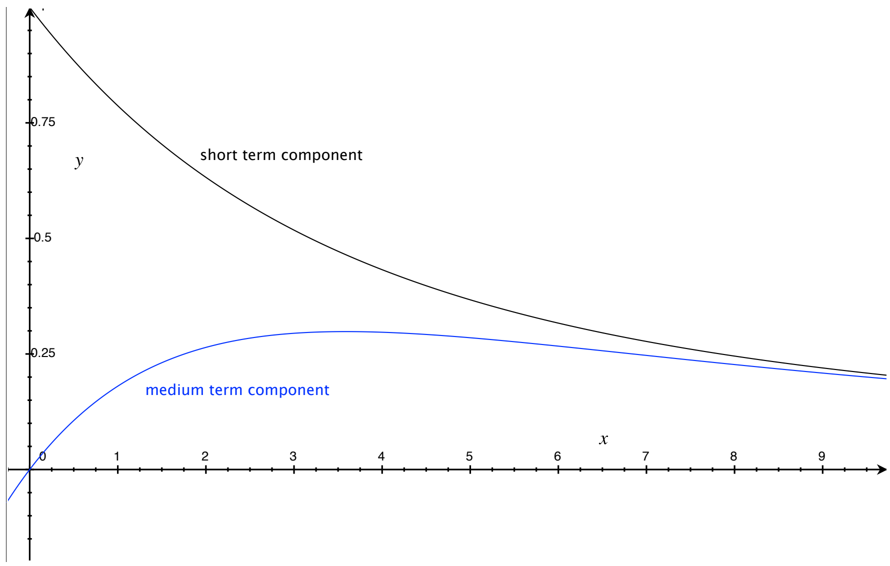


Figure 1:

## 4.3 Unique solution for $B_t$ and $G_t$ from Nelson-Seigal assumption

$$y_t\{\tau\} = -\frac{1}{\tau}B_t\{\tau\}\cdot X_t - \frac{1}{\tau}G_t\{\tau\} \tag{8}$$

$$y_t^N\{\tau\} = L_t^N + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{1}{\tau}G_t^N\{\tau\}$$

$$y_t^R\{\tau\} = L_t^R + \alpha^R S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \alpha^R C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) - \frac{1}{\tau}G_t^R\{\tau\}$$

(8) then implies

$$B_{1,t}^N\{\tau\} = B_{4,t}^R\{\tau\} = -\tau$$

$$B_{2,t}^N\{\tau\} = -\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right)$$

$$B_{2,t}^R\{\tau\} = -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right)$$

$$
\begin{aligned}
B^N_{3,t}\{\tau\} &= -\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\
B^R_{3,t}\{\tau\} &= -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\
B^R_{1,t}\{\tau\} &= B^N_{4,t}\{\tau\} = 0 \\
G^N_t\{0\} - G^N_t\{\tau\} &= \int_t^{t+\tau}\left(\rho_0 - B_s^{'}K_s^Q\theta_s^Q - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_s^{'}B_s^N B_s^{N'}\Sigma_s\right)_{\{j,j\}}\nu_s^{N(j)}\right)ds \\
G^N_t\{\tau\} &= \int_t^{t+\tau}\left(-\rho_0 + B_s^{'}K_s^Q\theta_s^Q + \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_s^{'}B_s^N B_s^{N'}\Sigma_s\right)_{\{j,j\}}\nu_s^{N(j)}\right)ds \\
G^R_t\{0\} - G^R_t\{\tau\} &= \int_t^{t+\tau}\left(\rho_0 - B_s^{R'}K_s^Q\theta_s^Q - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_s^{'}B_s^R B_s^{R'}\Sigma_s\right)_{\{j,j\}}\nu_s^{R(j)}\right)ds \\
G^R_t\{\tau\} &= \int_t^{t+\tau}\left(-\rho_0 + B_s^{R'}K_s^Q\theta_s^Q + \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_s^{'}B_s^R B_s^{R'}\Sigma_s\right)_{\{j,j\}}\nu_s^{R(j)}\right)ds \\
P_t\{0\} = 1 &\Rightarrow B_t\{0\} = G_t\{0\} = 0
\end{aligned}
$$

The above results combined with

$$
\begin{aligned}
\frac{d}{dt}B_t\{\tau\} &= \rho_1 + K_t^{Q'}B_t\{\tau\} - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t^{'}B_t\{\tau\}B_t\{\tau\}^{'}\Sigma_t\right)_{\{j,j\}}\left(\beta_t^{(j)}\right) \\
\frac{d}{dt}G_t\{\tau\} &= \rho_0 - B_t\{\tau\}^{'}K_t^Q\theta_t^Q - \frac{1}{2}\sum_{j=1}^{4}\left(\Sigma_t^{'}B_t\{\tau\}B_t\{\tau\}^{'}\Sigma_t\right)_{\{j,j\}}\nu_t^{(j)} \\
\beta_t^{(j)} &= 0 \quad \forall j \\
\nu_t^{(j)} &= 1 \quad \forall j
\end{aligned}
$$

imply

$$
\begin{aligned}
\rho_1^N &= \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad , \qquad \rho_1^R = \begin{pmatrix} 0 \\ \alpha^R \\ 0 \\ 1 \end{pmatrix} \\
K_t^{Q^{'}} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
K_t^Q &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda & -\lambda & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

We set $\rho_0^R = \rho_0^N = 0$ for identification.

### 4.4 Deriving the processes under the $P$ measure

Without loss of generality we can set $\theta^Q = 0$ because as we show below $X_t$ still follows a mean-reverting process under the physical probability measure, $P$.

$$\begin{aligned}
dX_t &= -K^Q X_t dt + \Sigma dW_t^Q \\
dX_t &= -K^Q X_t dt + \Sigma dW_t + \Sigma\Gamma dt \\
& \quad \text{To mantain the affined structure, we require} \\
\Gamma_t &= \underbrace{\begin{pmatrix} \gamma_0^{(1)} \\ \gamma_0^{(2)} \\ \gamma_0^{(3)} \\ \gamma_0^{(4)} \end{pmatrix}}_{\gamma_0} + \underbrace{\begin{pmatrix} \gamma_1^{(1)} \\ \gamma_1^{(2)} \\ \gamma_1^{(3)} \\ \gamma_1^{(4)} \end{pmatrix}}_{\gamma_1} X_t \\
\Rightarrow dX_t &= -K^Q X_t dt + \Sigma dW_t + \Sigma\Gamma_t \\
dX_t &= \Sigma\gamma_0 - (K^Q - \Sigma\gamma_1) X_t dt + \Sigma dW_t \\
\Rightarrow dX_t &= \underbrace{(K^Q - \Sigma\gamma_1)}_{K^P} \left( \underbrace{(K^Q - \Sigma\gamma_1)^{-1}\Sigma\gamma_0}_{\theta^P} - X_t \right) dt + \Sigma dW_t
\end{aligned}$$

Notice that $K^P$ and $\theta^P$ are unrestricted. Thus the model allows for plenty of flexibility. We then have

$$\begin{aligned}
B_t^N &= \begin{pmatrix} -(T-t) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ 0 \end{pmatrix} \\
B_t^R &= \begin{pmatrix} 0 \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda}\right) \\ -\alpha^R\left(\frac{1-e^{-\lambda(T-t)}}{\lambda} - e^{-\lambda(T-t)}(T-t)\right) \\ -(T-t) \end{pmatrix} \\
\frac{d}{dt} G_t^N &= -\frac{1}{2}\sum_{j=1}^{4}\left(\Sigma B_t^N B_t^{N'} \Sigma\right)_{\{j,j\}} \\
0 - G_t^N &= -\frac{1}{2}\int_t^T \left[\sum_{j=1}^{4}\left(\Sigma B_s^N B_s^{N'} \Sigma\right)_{\{j,j\}}\right] ds
\end{aligned}$$

$$
\begin{aligned}
G_t^N &= \frac{1}{2}\int_t^T \left[\sum_{j=1}^{4}\left(\Sigma B_s^N B_s^{N'}\Sigma\right)_{\{j,j\}}\right] ds \\
\frac{d}{dt}G_t^R &= -\frac{1}{2}\sum_{j=1}^{4}\left(\Sigma B_t^R B_t^{R'}\Sigma\right)_{\{j,j\}} \\
0 - G_t^R &= -\frac{1}{2}\int_t^T \left[\sum_{j=1}^{4}\left(\Sigma B_s^R B_s^{R'}\Sigma\right)_{\{j,j\}}\right] ds \\
G_t^R &= \frac{1}{2}\int_t^T \left[\sum_{j=1}^{4}\left(\Sigma B_s^R B_s^{R'}\Sigma\right)_{\{j,j\}}\right] ds
\end{aligned}
$$

As show in Section 2.3 of Christensen, Diebold and Rudebusch (2007) we need that $\Sigma$ is upper or lower-triangular for identification. Christensen, Diebold and Rudebusch (2007) also provide evidence that the out of sample forecasts are stronger when $\Sigma$ and $K^P$ are diagonal and (even though the in-sample fit is worse). It is unclear however whether the improvement in out-of-sample forecasts come from assuming that$\Sigma$ is diagonal or from instead that assuming $K^P$ is diagonal or both. Notice however that although a diagonal $\Sigma$ implies that the state variables are conditionally independent, the state variables will be correlated (unconditionally) if $K^P$ is kept unrestricted (as done in Christensen, Lopez and Rudebusch; 2007). We would need to compare the out-of-sample predictions of all four possible models ($\Sigma$ and $K^P$ diagonal, $\Sigma$ only diagonal, $\Sigma$ only unrestricted, $\Sigma$ and $K^P$ unrestricted) to really know which is best.

For now, in order to reduce some of the model's complexity, we will assume that $\Sigma$ is diagonal while keeping $K^P$ unrestricted. We then have

$$
\begin{aligned}
G_t^N &= \frac{\sigma_{11}^2}{2}\left(\int_t^T (T-s)^2 ds\right) + \frac{\sigma_{22}^2}{2}\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda}\right)^2 ds\right) \\
&+\frac{\sigma_{33}^2}{2}\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda} - e^{-\lambda(T-s)}(T-s)\right)^2 ds\right) \\
G_t^R &= \frac{\sigma_{22}^2}{2}(\alpha^R)^2\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda}\right)^2 ds\right) + \frac{\sigma_{33}^2}{2}(\alpha^R)^2\left(\int_t^T \left(\frac{1-e^{-\lambda(T-s)}}{\lambda} - e^{-\lambda(T-s)}(T-s)\right)^2 ds\right) \\
&+\frac{\sigma_{44}^2}{2}\left(\int_t^T (T-s)^2 ds\right)
\end{aligned}
$$

### 4.5 Kalman filter details

$$
\begin{aligned}
dX_t &= K^p\left(\theta^P - X_t\right)dt + \Sigma dW_t \\
dX_t + K^p X_t &= K^p\theta^P dt + \Sigma dW_t \\
exp\left(K^p(t-t_0)\right)dX_t + exp\left(K^p(t-t_0)\right)K^p X_t &= exp\left(K^p(t-t_0)\right)\left(K^p\theta^P dt + \Sigma dW_t\right)
\end{aligned}
$$

$$
\begin{aligned}
d\left(exp\left(K^p(t-t_0)\right)X_t\right) &= exp\left(K^p(t-t_0)\right)\left(K^p\theta^P dt + \Sigma dW_t\right)
\end{aligned}
$$

$$
\begin{aligned}
X_t &= exp\left(-K^p(t-t_0)\right)(K^p)^{-1}exp\left(K^p(t-t_0)\right)K^p\theta^P \\
&\quad + exp\left(-K^p(t-t_0)\right)\int_{t_0}^{t} exp\left(K^p(s-t_0)\right)\Sigma dW_s \\
&\quad + exp\left(-K^p(t-t_0)\right)Const \\
&\quad \text{So that at } t=t_0 \quad \text{we have } X_t = X_{t_0} \text{ we need} \\
X_{t_0} &= \theta^P + Const \\
\Rightarrow Const &= X_{t_0} - \theta^P \\
&\quad \text{Hence,}
\end{aligned}
$$

$$
\begin{aligned}
X_t &= exp\left(-K^p(t-t_0)\right)\left((K^p)^{-1}exp\left(K^p(t-t_0)\right)K^p\theta^P - \theta^P\right) \\
&\quad + exp\left(-K^p(t-t_0)\right)\int_{t_0}^{t} exp\left(K^p(s-t_0)\right)\Sigma dW_s \\
&\quad + exp\left(-K^p(t-t_0)\right)X_{t_0}
\end{aligned}
$$

$$
\begin{aligned}
X_t &= exp\left(-K^p(t-t_0)\right)\left((K^p)^{-1}exp\left(K^p(t-t_0)\right)K^p - I\right)\theta^P \\
&\quad + exp\left(-K^p(t-t_0)\right)\int_{t_0}^{t} exp\left(K^p(s-t_0)\right)\Sigma dW_s \\
&\quad + exp\left(-K^p(t-t_0)\right)X_{t_0}
\end{aligned}
$$

$$
E^P\left[X_t|\mathcal{F}_{t_0}\right] = \left(I - exp\left(-K^p(t-t_0)\right)\right)\theta^P + exp\left(-K^p(t-t_0)\right)X_{t_0}
$$

$$
\begin{aligned}
Var^P\left[X_t|\mathcal{F}_{t_0}\right] &= \int_{t_0}^{t} exp\left(-K^p(t-s)\right)\Sigma\Sigma' exp\left(-K^p(t-s)\right)' ds \\
&\quad \text{let } u = t-s \text{ using a change of variable:} \\
Var^P\left[X_t|\mathcal{F}_{t_0}\right] &= -\int_{t-t_0}^{0} exp\left(-K^p u\right)\Sigma\Sigma' exp\left(-K^p u\right)' du \\
Var^P\left[X_t|\mathcal{F}_{t_0}\right] &= \int_{0}^{t-t_0} exp\left(-K^p u\right)\Sigma\Sigma' exp\left(-K^p u\right)' du
\end{aligned}
$$

The discretized state equation is thus

$$
\begin{aligned}
X_i &= \left(I - exp\left(-K^p\Delta t_i\right)\right)\theta^P + exp\left(-K^p\Delta t_i\right)X_{t_i} + \eta_i \\
&\quad \text{where} \\
\Delta t_i &= \Delta t_i - t_{i-1} = \frac{1}{252} \quad \text{for daily data}
\end{aligned}
$$

$$
\eta_i \sim N(0,Q)
$$

$$Q = \int_0^{\Delta t_i} exp\left(-K^p u\right)\Sigma\Sigma^{'} exp\left(-K^p u\right)^{'} du$$

The measurement equations are

$$\begin{aligned}
y_t^N\{\tau\} &= A_0^N\{\tau\} + A_1^N\{\tau\}^{'} X_t + \epsilon_t^N \\
\\
y_t^R\{\tau\} &= A_0^R\{\tau\} + A_1^R\{\tau\}^{'} X_t + \epsilon_t^R \\
& \text{where} \\
\epsilon_t^N &\sim N(0, \phi_N^2\{\tau\}) \qquad \text{iid across } \tau \\
\epsilon_t^R &\sim N(0, \phi_R^2\{\tau\}) \qquad \text{iid across } \tau
\end{aligned}$$

$$\begin{aligned}
A_1^N\{\tau\} &= -\frac{1}{\tau}\begin{pmatrix} -\tau \\ -\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right) \\ -\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\ 0 \end{pmatrix} \\
A_0^N\{\tau\} &= -\frac{\sigma_{11}^2}{2\tau}\left(\int_t^{t+\tau}(t+\tau-s)^2 ds\right) - \frac{\sigma_{22}^2}{2\tau}\left(\int_t^{t+\tau}\left(\frac{1-e^{-\lambda(t+\tau-s)}}{\lambda}\right)^2 ds\right) \\
&\quad -\frac{\sigma_{33}^2}{2\tau}\left(\int_t^{t+\tau}\left(\frac{1-e^{-\lambda(t+\tau-s)}}{\lambda} - e^{-\lambda(t+\tau-s)}(t+\tau-s)\right)^2 ds\right) \\
A_1^R\{\tau\} &= -\frac{1}{\tau}\begin{pmatrix} 0 \\ -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda}\right) \\ -\alpha^R\left(\frac{1-e^{-\lambda\tau}}{\lambda} - e^{-\lambda\tau}\tau\right) \\ -\tau \end{pmatrix} \\
A_0^R\{\tau\} &= -\frac{\sigma_{22}^2}{2\tau}(\alpha^R)^2\left(\int_t^{t+\tau}\left(\frac{1-e^{-\lambda(t+\tau-s)}}{\lambda}\right)^2 ds\right) \\
&\quad -\frac{\sigma_{33}^2}{2\tau}(\alpha^R)^2\left(\int_t^{t+\tau}\left(\frac{1-e^{-\lambda(t+\tau-s)}}{\lambda} - e^{-\lambda(t+\tau-s)}(t+\tau-s)\right)^2 ds\right) \\
&\quad -\frac{\sigma_{44}^2}{2\tau}\left(\int_t^{t+\tau}(t+\tau-s)^2 ds\right)
\end{aligned}$$

Simplifying a bit

$$\begin{aligned}
A_0^N\{\tau\} &= -\frac{\sigma_{11}^2}{2\tau}\left(\frac{\tau^3}{3}\right) \\
&\quad -\frac{\sigma_{22}^2}{2\tau}\left(-\frac{-2\lambda\tau + e^{-2\lambda\tau} - 4e^{-\lambda\tau} + 3}{2\lambda^3}\right) \\
&\quad -\frac{\sigma_{33}^2}{2\tau}\left(\frac{e^{-2\lambda\tau}\left(8e^{\lambda\tau}(\lambda\tau+2) - 2\lambda\tau(\lambda\tau+3) + e^{2\lambda\tau}(4\lambda\tau-11) - 5\right)}{4\lambda^3}\right)
\end{aligned}$$

$$
\begin{aligned}
A_0^R\{\tau\} &= -\frac{\sigma_{22}^2}{2\tau}(\alpha^R)^2\left(-\frac{-2\lambda\tau+e^{-2\lambda\tau}-4e^{-\lambda\tau}+3}{2\lambda^3}\right) \\
&\quad -\frac{\sigma_{33}^2}{2\tau}(\alpha^R)^2\left(\frac{e^{-2\lambda\tau}\left(8e^{\lambda\tau}(\lambda\tau+2)-2\lambda\tau(\lambda\tau+3)+e^{2\lambda\tau}(4\lambda\tau-11)-5\right)}{4\lambda^3}\right) \\
&\quad -\frac{\sigma_{44}^2}{2\tau}\left(\frac{\tau^3}{3}\right)
\end{aligned}
$$

We will use several maturities ($T \in \{T_1, \ldots, T_M\}$ with $T_n = \frac{n}{\Delta_i}$; hence $\tau \in \{\tau_1, \ldots, \tau_M\}$ with $\tau_n = T_n - t$). We can stack the measurement equations

$$
\begin{aligned}
y_t &= A_0 + A_1 X_t + \epsilon_t \\
\epsilon_t^N &\sim N(0, \Phi) \\
A_1 &= \begin{pmatrix} A_1^N \\ A_1^R \end{pmatrix}, \qquad A_0 = \begin{pmatrix} A_0^N \\ A_0^R \end{pmatrix} \\
\Phi &= \begin{pmatrix} \phi_N^2 & 0 \\ 0 & \phi_R^2 \end{pmatrix} \\
A_0^N &= \begin{pmatrix} A_0^N\{\tau_1\} \\ \vdots \\ A_0^N\{\tau_M\} \end{pmatrix}, \qquad A_0^R = \begin{pmatrix} A_0^R\{\tau_1\} \\ \vdots \\ A_0^R\{\tau_M\} \end{pmatrix} \\
A_1^N &= \begin{pmatrix} A_1^N\{\tau_1\}' \\ \vdots \\ A_1^N\{\tau_M\}' \end{pmatrix}, \qquad A_1^R = \begin{pmatrix} A_1^R\{\tau_1\}' \\ \vdots \\ A_1^R\{\tau_M\}' \end{pmatrix} \\
\phi_N^2 &= \begin{pmatrix} \phi_N^2\{\tau_1\} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_N^2\{\tau_M\} \end{pmatrix}, \qquad \phi_R^2 = \begin{pmatrix} \phi_R^2\{\tau_1\} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_R^2\{\tau_M\} \end{pmatrix}
\end{aligned}
$$

### 4.5.1 Kalman gain

Let us first derive the prediction step

$$
\begin{aligned}
\hat{X}_{k|k-1} &= (I - exp(-K^p\Delta))\,\theta^P + exp(-K^p\Delta)\,\hat{X}_{k-1|k-1} \\
X_k - \hat{X}_{k|k-1} &= exp(-K^p\Delta)\left(X_{k-1} - \hat{X}_{k-1|k-1}\right) + \eta_k
\end{aligned}
$$

$$\Rightarrow \hat{V}_{k|k-1} = exp\left(-K^p\Delta\right)\hat{V}_{k-1|k-1}exp\left(-K^p\Delta\right)^{'} + Q_k$$

The update steps are derived below. $G_k$ minimizes the mean square posteriori error $\left(E\left[||X_k - \hat{X}_{k|k}||^2\right]\right)$ given$\hat{X}_{k|k} = \hat{X}_{k|k-1} + G_k\varepsilon_k$. Let $\hat{V}_{k|k} = cov\left(X_k - \hat{X}_{k|k}\right)$ and $\hat{V}_{k|k-1} = cov\left(X_k - \hat{X}_{k|k-1}\right)$

$$\begin{aligned}
\hat{V}_{k|k} &= cov\left(X_k - \hat{X}_{k|k}\right) \\
&= cov\left(X_k - \left(\hat{X}_{k|k-1} + G_k\varepsilon_k\right)\right) \\
&= cov\left(X_k - \left(\hat{X}_{k|k-1} + G_k\left(y_k - A_0 - A_1\hat{X}_{k|k-1}\right)\right)\right) \\
&= cov\left(X_k - \left(\hat{X}_{k|k-1} + G_k\left(A_0 + A_1X_k + \epsilon_t - A_0 - A_1\hat{X}_{k|k-1}\right)\right)\right) \\
&= cov\left(\left(I - G_kA_1\right)\left(X_k - \hat{X}_{k|k-1}\right) - G_k\epsilon_t\right) \\
&= \left(I - G_kA_1\right)cov\left(X_k - \hat{X}_{k|k-1}\right)\left(I - G_kA_1\right)^{'} + G_kcov\left(\epsilon_t\right)G_k^{'} \\
\hat{V}_{k|k} &= \left(I - G_kA_1\right)\hat{V}_{k|k-1}\left(I - G_kA_1\right)^{'} + G_k\Phi G_k^{'} \\
\hat{V}_{k|k} &= \left(\hat{V}_{k|k-1} - G_kA_1\hat{V}_{k|k-1}\right)\left(I - G_kA_1\right)^{'} + G_k\Phi G_k^{'} \\
\hat{V}_{k|k} &= \hat{V}_{k|k-1}\left(I - G_kA_1\right)^{'} - G_kA_1\hat{V}_{k|k-1} + G_kS_kG_k^{'}
\end{aligned}$$

$$\begin{aligned}
E\left[||\hat{X}_k - \hat{X}_{k|k}||^2\right] &= tr\left(\hat{V}_{k|k}\right) \\
& \quad \text{FOC wrt } G_k: \\
0 &= -\hat{V}_{k|k-1}\left(A_1\right)^{'} - \left(A_1\hat{V}_{k|k-1}\right)^{'} + 2G_kS_k \\
& \quad \text{since } \hat{V}_{k|k-1} \text{ is symmetric} \\
0 &= 2G_kS_k - 2(\hat{V}_{k|k-1})^{'}A_1^{'} \\
\Rightarrow G_k &= (\hat{V}_{k|k-1})^{'}A_1^{'}(S_k)^{-1} = \hat{V}_{k|k-1}A_1^{'}(S_k)^{-1} \\
\hat{V}_{k|k} &= \hat{V}_{k|k-1}\left(I - \hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}A_1\right)^{'} - \hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}A_1\hat{V}_{k|k-1} \\
& \quad +\hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}S_k\left(\hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}\right)^{'}
\end{aligned}$$

$$\hat{V}_{k|k} = \hat{V}_{k|k-1} - 2\hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}A_1\hat{V}_{k|k-1} + \hat{V}_{k|k-1}A_1^{'}\left(\hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}\right)^{'}$$

$$\hat{V}_{k|k} = \hat{V}_{k|k-1} - 2\hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}A_1\hat{V}_{k|k-1} + \hat{V}_{k|k-1}A_1^{'}\left((S_k)^{-1}\right)^{'}A_1\hat{V}_{k|k-1}^{'}$$

$$\hat{V}_{k|k} = \hat{V}_{k|k-1} - \hat{V}_{k|k-1}A_1^{'}(S_k)^{-1}A_1\hat{V}_{k|k-1}$$

$$\Rightarrow \hat{V}_{k|k} = \hat{V}_{k|k-1} - G_kA_1\hat{V}_{k|k-1}$$

$$\varepsilon_k = y_k - A_0 - A_1\hat{X}_{k|k-1} \qquad \text{innovation}$$

$$\varepsilon_k = \epsilon_k + A_1\left(X_k - \hat{X}_{k|k-1}\right)$$

$$S_k = cov(\varepsilon_k) = \Phi + A_1\hat{V}_{k|k-1}A_1$$

### 4.5.2 Initializing error variance

We can set $\hat{V}_{0|0}$ to the stationary error variance or the unconditional error variance. Below we derive the stationary error variance

$$\hat{V}_{k|k} = \hat{V}_{k|k-1} - G_kA_1\hat{V}_{k|k-1}$$

$$\hat{V}_{k|k-1} = exp\left(-K^p\Delta\right)\hat{V}_{k-1|k-1}exp\left(-K^p\Delta\right)^{'} + Q$$

Combining the prediction and updating equtions:

$$\hat{V}_{k|k-1} = exp\left(-K^p\Delta\right)\left(\hat{V}_{k-1|k-2} - G_{k-1}A_1\hat{V}_{k-1|k-2}\right)exp\left(-K^p\Delta\right)^{'} + Q$$

$$\hat{V}_{k|k-1} = exp\left(-K^p\Delta\right)\left(\hat{V}_{k-1|k-2} - \left(\hat{V}_{k-1|k-2}A_1^{'}(S_{k-1})^{-1}\right)A_1\hat{V}_{k-1|k-2}\right)exp\left(-K^p\Delta\right)^{'} + Q$$

$$\hat{V}_{k|k-1} = exp\left(-K^p\Delta\right)\left(\hat{V}_{k-1|k-2} - \left(\hat{V}_{k-1|k-2}A_1^{'}\left(\Phi + A_1\hat{V}_{k-1|k-2}A_1^{'}\right)^{-1}\right)A_1\hat{V}_{k-1|k-2}\right)exp\left(-K^p\Delta\right)^{'} + Q$$

$$\begin{aligned}\hat{V}_{k|k-1} = {} & exp\left(-K^p\Delta\right)\hat{V}_{k-1|k-2}exp\left(-K^p\Delta\right)^{'} \\ & -exp\left(-K^p\Delta\right)\hat{V}_{k-1|k-2}A_1^{'}\left(\Phi + A_1\hat{V}_{k-1|k-2}A_1^{'}\right)^{-1}A_1\hat{V}_{k-1|k-2}exp\left(-K^p\Delta\right)^{'} + Q\end{aligned}$$

The stationary error variance solve the following Riccati equation:

$$
\begin{aligned}
V &= exp\left(-K^{p}\Delta\right)Vexp\left(-K^{p}\Delta\right)^{'} \\
&\quad -exp\left(-K^{p}\Delta\right)VA_{1}^{'}\left(\Phi+A_{1}VA_{1}^{'}\right)^{-1}A_{1}Vexp\left(-K^{p}\Delta\right)^{'}+Q
\end{aligned}
$$

Below we derive the unconditional error variance[9]

$$
\begin{aligned}
\hat{V}_{k|k-1} &= exp\left(-K^{p}\Delta\right)\hat{V}_{k-1|k-1}exp\left(-K^{p}\Delta\right)^{'}+Q \\
&\quad \text{The unconditional variance solves the discrete Lyapunov equation below} \\
V &= exp\left(-K^{p}\Delta\right)Vexp\left(-K^{p}\Delta\right)^{'}+Q \\
&\quad \text{Equivalently we can write:} \\
vec\left(V\right) &= \left(exp\left(-K^{p}\Delta\right)\otimes exp\left(-K^{p}\Delta\right)^{'}\right)vec\left(V\right)+vec\left(Q\right) \\
\Rightarrow vec\left(V\right) &= \left(I-exp\left(-K^{p}\Delta\right)\otimes exp\left(-K^{p}\Delta\right)\right)^{-1}vec\left(Q\right)
\end{aligned}
$$

equivalently we can write

$$
\begin{aligned}
\hat{V}_{k|k-1} &= exp\left(-K^{p}\Delta\right)\hat{V}_{k-1|k-1}exp\left(-K^{p}\Delta\right)^{'}+Q \\
V=V_{\infty} &= \left(\sum_{n=0}^{\infty}exp\left(-K^{p}\Delta n\right)Qexp\left(-K^{p}\Delta n\right)^{'}\right)\Delta \\
\Rightarrow V &= \int_{0}^{\infty}exp\left(K^{p}s\right)\Sigma\Sigma^{'}exp\left(K^{p}s\right)^{'}ds \\
&\quad \text{Thus, the unconditional variance solves the continuous Lyapunov equation} \\
0 &= e^{-Kt}V+Ve^{-K't}+\Sigma\Sigma^{'} \\
&\quad \text{Equivalently we can write:} \\
vec\left(V\right) &= \left(I\otimes exp\left(-K^{p}\Delta\right)+exp\left(-K^{p}\Delta\right)\otimes I\right)^{-1}\left(-vec\left(\Sigma\Sigma^{'}\right)\right)
\end{aligned}
$$

Christensen, Lopez and Rudebusch (2010) initialize the error variance using the unconditional error variance.

### 4.5.3 On Lyapunov equations

$$
\begin{aligned}
Q\{h\} &= \int_{0}^{h}e^{-Kt}Se^{-K't}dt \\
&\quad \text{Let} \\
V &= \lim_{h\to\infty}Q\{h\} \\
\Rightarrow V &= \int_{0}^{\infty}e^{-Kt}Se^{-K't}dt
\end{aligned}
$$

[9] https://en.wikipedia.org/wiki/Lyapunov_equation

Thus $V$ is the unique solution to a continuous Lyapunov equation of the form

$$e^{-Kt}V + Ve^{-K't} + S \quad = \quad 0$$

where **$S$ is symmetric** and $K$ and $S$ are stable.

That is, the **real component of the eigenvalues of $K$ and $S$ are positive**.

Indeed, by vectorizing we find

$$\begin{aligned}
e^{-Kt}V + Ve^{-K't} &= -S \\
vec(e^{-Kt}V + Ve^{-K't}) &= -vec(S) \\
\left(I \otimes e^{-Kt} + e^{-Kt} \otimes I\right) vec(V) &= -vec(S) \\
\Rightarrow vec(V) = vec\left(\int_0^\infty e^{-Kt}Se^{-K't}dt\right) &= \left(I \otimes e^{-Kt} + e^{-Kt} \otimes I\right)^{-1}(-vec(S)) \\
\Rightarrow vec(Q(\infty)) &= -\left(I \otimes e^{-Kt} + e^{-Kt} \otimes I\right)^{-1} vec(S)
\end{aligned}$$

We can also instead find a "discrete" solution

$$\begin{aligned}
Q\{\Delta\} &= \int_0^\Delta e^{-Kt}Se^{-K't}dt = e^{-K\Delta}Se^{-K'\Delta}\Delta \\
V &= \int_0^\infty e^{-Kt}Se^{-K't}dt \\
V &= \left(\sum_{m=1}^\infty e^{-K\Delta m}Se^{-K'\Delta m}\right)\Delta = \left(\sum_{m=0}^\infty e^{-K\Delta(m-1)}Q\{\Delta\}e^{-K'\Delta(m-1)}\right) = \sum_{n=0}^\infty e^{-K\Delta n}Q\{\Delta\}e^{-K'\Delta n} \\
\Rightarrow V &= \sum_{n=0}^\infty e^{-K\Delta n}Q\{\Delta\}e^{-K'\Delta n}
\end{aligned}$$

Thus $V$ is the unique solution to a discrete Lyapunov equation of the form

$$V \quad = \quad e^{-K\Delta}Ve^{-K'\Delta} + Q\{\Delta\}$$

where **$S$ is symmetric** and $K$ and $S$ are stable.

That is, the **real component of the eigenvalues of $K$ and $S$ are positive**.

Indeed, by vectorizing we find

$$\begin{aligned}
vec\,(V) &= \left(e^{-K\Delta} \otimes e^{-K\Delta}\right) vec\,(V) + vec\,(Q\{\Delta\}) \\
\Rightarrow vec\,(V) &= \left(I - e^{-K\Delta} \otimes e^{-K\Delta}\right)^{-1} vec\,(Q\{\Delta\})
\end{aligned}$$

See vectorizing rules in https://en.wikipedia.org/wiki/Vectorization_(mathematics)

### 4.5.4 Kalman Smoother and EM algorithm

This derivation is partly obtained from the "Time Series" lecture notes of Francis Diebold[10]. After running the Kalman Filter an storing $X_{t,t}$, $V_{t,t}$, $V_{t+1,t}$, $G$ (the Kalman gain) we then proceed as follows.

The smoothed states are

$$X_{t,T} = X_{t,t} + J_t \left( X_{t+1,T} - X_{t+1,t} \right)$$

$$V_{t,T} = V_{t,t} + J_t \left( V_{t+1,T} - V_{t+1,t} \right) J_t^{'}$$

where

$$J_t = V_{t,t} U_1^{'} \left( V_{t+1,t} \right)^{-1}$$
$$J_0 = V_0 U_1^{'} \left( V_{1,0} \right)^{-1}$$
$$X_{0,T} = X_0 + J_0 \left( X_{1,T} - X_{1,0} \right)$$

$$E\left[ X_{t,t} X_{t,t}^{'} \middle| X^{(T)} \right] = Cov\left[ X_{t,t}, X_{t,t} \middle| X^{(T)} \right] + X_{t,T} X_{t,T}^{'} = V_{t,T} + X_{t,T} X_{t,T}^{'}$$

$$E\left[ X_{t,t} X_{t-1,t-1}^{'} \middle| X^{(T)} \right] = Cov\left[ X_{t,t}, X_{t-1,t-1} \middle| X^{(T)} \right] + X_{t,T} X_{t-1,T}^{'} = V_{(t,t-1),T} + X_{t,T} X_{t-1,T}^{'}$$

to find the smoothed predictive cov. matrix, $V_{(t,t-1),T}$, we use

$$V_{(T,T-1),T} = (I - G_T A_1) U_1 V_{T-1,T-1}$$

$$V_{(t-1,t-2),T} = V_{t-1,t-1} J_{t-2}^{'} + J_{t-1} \left( V_{(t,t-1),T} - U_1 V_{t-1,t-1} \right) J_{t-2}^{'}$$

[11]

## 4.6 Alternative estimation methods

We could instead estimate $X_t$ via PCA and then use

### 4.6.1 MLE

We would first stack $X_t$ and $y_t$

$$Y_t = \begin{pmatrix} X_t \\ y_t \end{pmatrix}$$

$$Y_t = B_0 + B_1 Y_{t-1} + \mu_t$$

$$\mu_t \sim N\left( 0, \begin{pmatrix} Q & 0 \\ 0 & \Phi \end{pmatrix} \right)$$

[10] http://www.ssc.upenn.edu/~fdiebold/Teaching706/TimeSeriesSlides.pdf

[11] Notice that $det(exp(A)) = exp(tr(A)) \Rightarrow \frac{1}{2} ln \left( |\Sigma_0| \right) = \frac{1}{2} tr \left( ln \left( \Sigma_0 \right) \right)$. Recall also that $\frac{\partial}{\partial A} tr(AB) = B^{'}$ (http://www.cs.berkeley.edu/~jduchi/projects/matrix_prop.pdf)

we would then use MLE. The standard errors may need to be corrected for errors coming from the PCA. This can be done using a Bootstrap approach but would be very time consuming.

### 4.7 Finding moments of inflation

Below we use the forward approach[12]

$$
\begin{aligned}
ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right) &= \int_t^{t+\tau} \left(\rho^N - \rho^R\right)' X_s ds \\
\frac{\partial}{\partial\tau} ln\left(\frac{\Pi_{t+\tau}}{\Pi_t}\right) &= \left(\rho^N - \rho^R\right)' X_{t+\tau} \\
Z_{t,t+\tau} &\equiv \begin{pmatrix} X_{t+\tau} \\ ln\left(\Pi_{t+\tau}/\Pi_t\right) \end{pmatrix} \qquad \text{the augmented state vector} \\
dZ_{t,t+\tau} &= \underbrace{\begin{pmatrix} K^P & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \theta^P \\ 0 \end{pmatrix}}_{K} dt - \underbrace{\begin{pmatrix} K^p & 0 \\ -\left(\rho^N - \rho^R\right)' & 0 \end{pmatrix}}_{-\theta} Z_{t,t+\tau} dt + \underbrace{\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}}_{\bar{\Sigma}} \underbrace{\begin{pmatrix} dW_{t+\tau} \\ 0 \end{pmatrix}}_{d\bar{W}_{t+\tau}} \\
Z_{t,t+\tau} &= Z_{t,t} + \int_t^{t+\tau} K ds + \int_t^{t+\tau} \theta Z_{t,s} ds + \int_t^{t+\tau} \bar{\Sigma} d\bar{W}_s \\
m_{t,t+\tau} = E_t\left[Z_{t,t+\tau}\right] &= Z_{t,t} + \int_t^{t+\tau} K ds + E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right] \equiv m\left\{\tau\right\} \\
\frac{\partial}{\partial\tau} m\left\{\tau\right\} &= K + \theta m\left\{\tau\right\} \\
m_{t,t} = m\left\{0\right\} = &= Z_{t,t} = \begin{pmatrix} X_t \\ 0 \end{pmatrix} \\
v_{t,t+\tau} = Var_t\left[Z_{t,t+\tau}\right] = v\left\{\tau\right\} &= E_t\left[\left(\int_t^{t+\tau} \theta Z_{t,s} ds - E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right]\right)\left(\int_t^{t+\tau} \theta Z_{t,s} ds - E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right]\right)'\right] \\
&\quad + \int_t^{t+\tau} \bar{\Sigma}\bar{\Sigma}' ds \\
\frac{\partial}{\partial\tau} v\left\{\tau\right\} &= E_t\left[\left(\theta Z_{t,t+\tau} - E_t\left[\theta Z_{t,t+\tau} ds\right]\right)\left(\int_t^{t+\tau} \theta Z_{t,s} ds - E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right]\right)'\right] \\
&\quad + E_t\left[\left(\int_t^{t+\tau} \theta Z_{t,s} ds - E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right]\right)\left(\theta Z_{t,t+\tau} - E_t\left[\theta Z_{t,t+\tau} ds\right]\right)'\right] \\
&\quad + \bar{\Sigma}\bar{\Sigma}' \\
&\quad \text{using:} \\
Z_{t,t+\tau} - E_t\left[Z_{t,t+\tau}\right] &= \int_t^{t+\tau} \theta Z_{t,s} ds - E_t\left[\int_t^{t+\tau} \theta Z_{t,s} ds\right] + \int_t^{t+\tau} \bar{\Sigma} d\bar{W}_s
\end{aligned}
$$

yields:

[12] This is partly reproduced from Christensen, Lopez, Rudebusch (2013)

$$\begin{aligned}\frac{\partial}{\partial\tau}v\left\{\tau\right\} &= E_t\left[\left(\theta\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]+\int_t^{t+\tau}\bar{\Sigma}d\bar{W}_s\right)\right)\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]\right)^{'}\right]\\ &+E_t\left[\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]\right)\left(\theta\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]+\int_t^{t+\tau}\bar{\Sigma}d\bar{W}_s\right)\right)^{'}\right]\\ &+\bar{\Sigma}\bar{\Sigma}^{'}\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\tau}v\left\{\tau\right\} &= \theta v\left\{\tau\right\}-\theta\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds+E_t\left[\left(\theta\int_t^{t+\tau}\bar{\Sigma}d\bar{W}_s\right)\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]\right)^{'}\right]\\ &+v\left\{\tau\right\}\theta^{'}-\left(\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds\right)\theta^{'}+E_t\left[\left(\int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right]\right)\left(\theta\int_t^{t+\tau}\bar{\Sigma}d\bar{W}_s\right)^{'}\right]\\ &+\bar{\Sigma}\bar{\Sigma}^{'}\end{aligned}$$

$$\begin{aligned} & \text{using:}\\ \int_t^{t+\tau}\theta Z_{t,s}ds-E_t\left[\int_t^{t+\tau}\theta Z_{t,s}ds\right] &= Z_{t,t+\tau}-E_t\left[Z_{t,t+\tau}\right]-\int_t^{t+\tau}\bar{\Sigma}d\bar{W}_s\\ & \text{yields:}\\ \frac{\partial}{\partial\tau}v\left\{\tau\right\} &= \theta v\left\{\tau\right\}-\theta\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds+E_t\left[\left(\theta\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds\right)\right]\\ &+v\left\{\tau\right\}\theta^{'}-\left(\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds\right)\theta^{'}+E_t\left[\left(\theta\int_t^{t+\tau}\bar{\Sigma}\bar{\Sigma}^{'}ds\right)\right]\\ &+\bar{\Sigma}\bar{\Sigma}^{'}\end{aligned}$$

$$\frac{\partial}{\partial\tau}v\left\{\tau\right\} = \theta v\left\{\tau\right\}+v\left\{\tau\right\}\theta^{'}+\bar{\Sigma}\bar{\Sigma}^{'}$$

$$\begin{aligned}v_{t,t}=v\left\{0\right\} &= 0_{n\times n}\quad\text{where } n \text{ is the size of the augmented state vector}\\ & \text{It will be convenient to first vectorize } v_{t,t+\tau}\\ \frac{\partial}{\partial\tau}vec\left(v\left\{\tau\right\}\right) &= vec\left(I_{n\times n}\theta v\left\{\tau\right\}\right)+vec\left(v\left\{\tau\right\}\theta^{'}\right)+vec\left(I_{n\times n}\bar{\Sigma}\bar{\Sigma}^{'}\right)\\ \frac{\partial}{\partial\tau}vec\left(v\left\{\tau\right\}\right) &= \left(I_{n\times n}\otimes\theta\right)vec\left(v\left\{\tau\right\}\right)+\left(\theta\otimes I_{n\times n}\right)vec\left(v\left\{\tau\right\}\right)+\left(I_{n\times n}\otimes\bar{\Sigma}\right)vec\left(\bar{\Sigma}^{'}\right)\end{aligned}$$

Notice that $v\left\{\tau\right\}$ does not depend on $t$ because $v_{t,t}=v\left\{0\right\}=0_{n\times n}$.

We now have

$$\begin{aligned}-\left(\rho^N-\rho^R\right)^{'}\int_t^{t+\tau}X_sds &\sim \mathcal{N}\left(-m_{t,t+\tau},v\left\{\tau\right\}\right)\\ E_t\left[exp\left(-\left(\rho^N-\rho^R\right)^{'}\int_t^{t+\tau}X_sds\right)\right] &= exp\left(-m_{t,t+\tau}^{(n)}+\frac{1}{2}v^{(n,n)}\left\{\tau\right\}\right)\\ \Rightarrow\pi_t^e\left\{\tau\right\} &= -\frac{1}{\tau}\left(-m_{t,t+\tau}^{(n)}+\frac{1}{2}v^{(n,n)}\left\{\tau\right\}\right)\end{aligned}$$

Let us verify that the solution is indeed correct

$$
\begin{aligned}
\pi_t^e\{\tau\} &= -\frac{1}{\tau}ln\left(E_t\left[exp\left(-\int_t^{t+\tau}(r_s^N - r_s^R)ds\right)\right]\right)\\
&= -\frac{1}{\tau}ln\left(E_t\left[exp\left(-(\rho^N-\rho^R)'\int_t^{t+\tau}X_s ds\right)\right]\right)\\
&\quad \text{Using L'Hopital's rule:}\\
\lim_{\tau\downarrow 0}\pi_t^e\{\tau\} &= \pi_t^e\{0\} = (\rho^N-\rho^R)' X_t\\
&\quad \text{On the other hand, using L'Hopital's rule:}\\
\lim_{\tau\downarrow 0}\pi_t^e\{\tau\} &= -\left(-\frac{\partial}{\partial\tau}m_{t,t+\tau}^{(n)} + \frac{1}{2}\frac{\partial}{\partial\tau}v^{(n,n)}\{\tau\}\right)\Big|_{\tau=0}\\
\lim_{\tau\downarrow 0}\pi_t^e\{\tau\} &= \frac{\partial}{\partial\tau}m_{t,t+\tau}^{(n)} - \frac{1}{2}\frac{\partial}{\partial\tau}v^{(n,n)}\{\tau\}\Big|_{\tau=0}\\
\lim_{\tau\downarrow 0}\pi_t^e\{\tau\} &= K^{(n,n)} + (\theta m\{\tau\})^{(n,n)} - 0\\
\lim_{\tau\downarrow 0}\pi_t^e\{\tau\} &= 0 + (\theta m\{0\})^{(n,n)} = (\rho^N-\rho^R)' X_t \qquad \text{QED}
\end{aligned}
$$

Analytical solution:

$$
\begin{aligned}
\frac{\partial}{\partial\tau}m\{\tau\} &= K + \theta m\{\tau\}\\
\frac{\partial}{\partial\tau}m\{\tau\} - \theta m\{\tau\} &= K\\
exp(-\theta\tau)\frac{\partial}{\partial\tau}m\{\tau\} - exp(-\theta\tau)\,\theta m\{\tau\} &= exp(-\theta\tau)K\\
\frac{\partial}{\partial\tau}\left(exp(-\theta\tau)\,m\{\tau\}\right) &= exp(-\theta\tau)K\\
exp(-\theta\tau)\,m\{\tau\} &= \int exp(-\theta u)K du
\end{aligned}
$$

Since $\theta$ and $K$ are singular, we will rely on numerical methods to solve the ODE's rather than look for analytical solutions.

## 4.8 Currency exchange rates implications

Suppose we have two countries (home, $h$ and foreign, $f$). For simplicity we assume that the $X_t^h$ and $X_t^f$ are conditionally independent. The exchange rate is

$$
\begin{aligned}
e_t^{h,f} &= \frac{\Pi_t^h}{\Pi_t^f}\\
e_t^{h,f} &= exp\left(\int_0^t\left((r_s^{N,h}-r_s^{N,f}) - (r_s^{R,h}-r_s^{R,f})\right)ds\right)
\end{aligned}
$$

$$\frac{de_t^{h,f}}{e_t^{h,f}} = \left((r_t^{N,h} - r_t^{N,f}) - (r_t^{R,h} - r_t^{R,f})\right) dt$$

$$dln\left(e_t^{h,f}\right) = \left((r_t^{N,h} - r_t^{N,f}) - (r_t^{R,h} - r_t^{R,f})\right) dt$$

We can then compare the model's implied exchange rates to realized exchange rates. We can also calculate model's implied expected exchange rate over any horizon

$$m_\tau \equiv m_{0,\tau}^{(n)}$$

$$v_\tau \equiv v^{(n,n)}\{\tau\}$$

$$E_t\left[e_{t+\tau}^{h,f}\right] = exp\left(m_\tau^h - m_\tau^f - \frac{1}{2}\left(v_\tau^h - v_\tau^f\right)\right)$$