Appendix - Bond Liquidity Premia

Jean-Sébastien Fontaine René Garcia

Bank of Canada *EDHEC Business School*

A Arbitrage-Free Nelson-Siegel Term Structure Models

This section follows Christensen et al. (2007) and provides a description of the term structure model. The $k = 3$ term structure factors are stacked in the vector F_t. Its dynamics under the risk-neutral measure $\mathbb{Q}$ is described by the stochastic differential equation

$$dF_t^{\mathbb{Q}} = K^{\mathbb{Q}}(\theta^{\mathbb{Q}} - F_t) + \Sigma dW_t^{\mathbb{Q}}, \tag{1}$$

where dW_t is a standard Brownian motion process. Combined with the assumption that the short rate is affine in all three factors, this leads to the usual affine solution for discount bond yields. In this context, CDR show that if the short rate is defined as $r_t = F_{1,t} + F_{2,t}$ and if the mean-reversion matrix $K^{\mathbb{Q}}$ is restricted to

$$K^{\mathbb{Q}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix}, \tag{2}$$

then the absence of arbitrage opportunity implies the discount yield function,

$$y(F_t, m) = a(m) + F_{1,t}b_1(m) + F_{2,t}b_2(m) + F_{3,t}b_3(m). \tag{3}$$

Figure 1 displays the loadings. These are given by,

$$\begin{aligned} b_1(m) &= 1, \\ b_2(m) &= \left(\frac{1 - \exp(-m\lambda)}{m\lambda} \right), \\ b_3(m) &= \left(\frac{1 - \exp(-m\lambda)}{m\lambda} - \exp(-m\lambda) \right), \end{aligned} \tag{4}$$

where $m \geq 0$ is the length of time until maturity. CDR show that we are free to choose the drift term under for the dynamics under the historical measure, $\mathbb{P}$,

$$F_{t+1}^{\mathbb{P}} = K^{\mathbb{P}}(\theta^{\mathbb{P}} - F_t) + \Sigma dW_t^{\mathbb{P}}. \tag{5}$$

This relies on a linear specification of the prices of risk. Any choice of $\theta^{\mathbb{P}}$ and $K^{\mathbb{P}}$ matrix pins downs all the price of risk parameters. Note that the first factor has a unit root under the risk-neutral density. Then, as discussed in CDR, we have that $a(m) \to -\infty$ when $m \to \infty$. However, $a(m)$ is small for most maturities. In particular, at parameter estimates from Table 3 below, we have that $a \approx 0$ for short maturities and $a(m) \approx -1\%$ at a maturity of 30 years. In contrast, we only consider maturities of 10 years or less. If we assume $\theta_{\mathbb{Q}} = 0$ to identify the mean of F_t under $\mathbb{Q}$ then $a(m)$ is given by

$$\begin{aligned} a(m) = & -\sigma_{11}^2 \frac{m^2}{6} - (\sigma_{21}^2 + \sigma_{22}^2) \left[\frac{1}{2\lambda^2} - \frac{1 - e^{-m\lambda}}{m\lambda^3} + \frac{1 - e^{-2m\lambda}}{4m\lambda^3} \right] \\ & - (\sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2) \left[\frac{1}{2\lambda^2} + \frac{e^{-m\lambda}}{\lambda^2} - \frac{me^{-2m\lambda}}{4\lambda} - \frac{3e^{-2m\lambda}}{4\lambda^2} - \frac{2(1 - e^{-m\lambda})}{m\lambda^3} + \frac{5(1 - e^{-2m\lambda})}{8m\lambda^3} \right] \\ & - (\sigma_{11}\sigma_{21}) \left[\frac{m}{2\lambda} + \frac{e^{-m\lambda}}{\lambda^2} - \frac{1 - e^{-m\lambda}}{m\lambda^3} \right] - (\sigma_{11}\sigma_{31}) \left[\frac{3e^{-m\lambda}}{\lambda^2} + \frac{m}{-2\lambda} + \frac{me^{-m\lambda}}{\lambda} \right] \\ & - (\sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32}) \times \left[\frac{1}{\lambda^2} + \frac{e^{-m\lambda}}{\lambda^2} - \frac{e^{-2m\lambda}}{\lambda^2} - \frac{3(1 - e^{-m\lambda})}{m\lambda^3} + \frac{3(1 - e^{-2m\lambda})}{4m\lambda^3} \right]. \end{aligned} \tag{6}$$

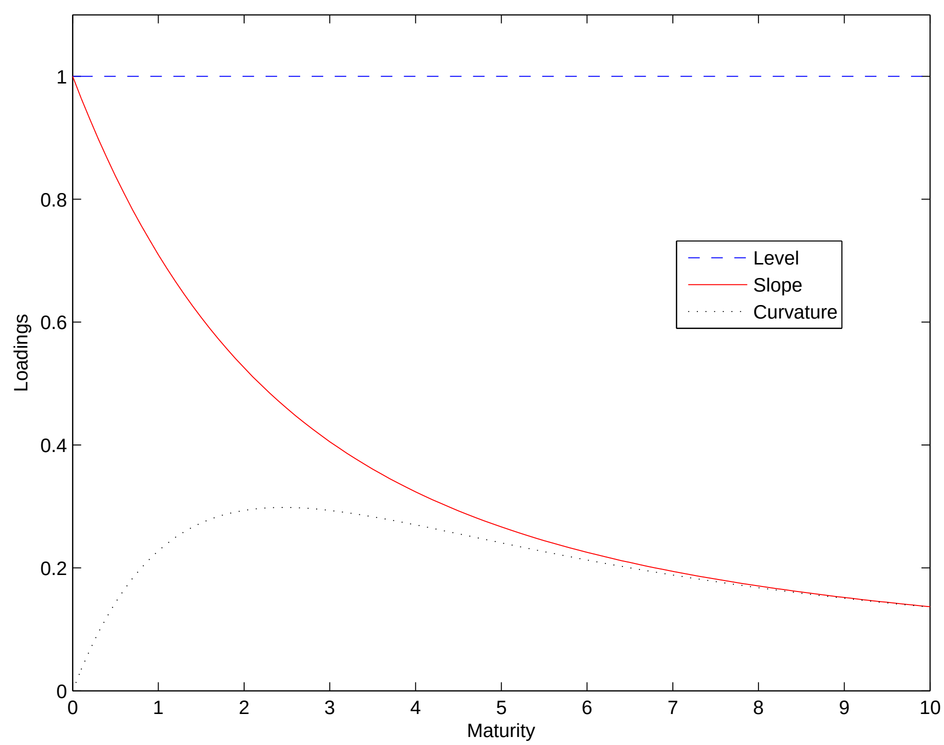

Figure 1: **Factor Loadings** Estimated level, slope and curvature factor loadings from the term structure model with liquidity. End-of-month data from CRSP (1985:12-2007:12).

B State-Space Representation And Likelihood Function

The term structure factors have the following discretized state equation,

$$(F_t - \bar{F}) = \Phi_f(F_{t-1} - \bar{F}) + \Omega_f \epsilon_t, \tag{7}$$

where ϵ_t is standard Gaussian, the autoregressive matrix Φ and the covariance matrix Γ are

$$\Phi_f = \exp\left(-K\frac{1}{12}\right) \qquad \Omega_f = \int_0^{\frac{1}{12}} e^{-Ks}\Sigma\Sigma^T e^{-Ks} ds. \tag{8}$$

The transition equation for L_t is given by

$$(L_t - \bar{L}) = \phi_l(L_{t-1} - \bar{L}) + \sigma_l \epsilon_t^l, \tag{9}$$

where ϵ_t^l is standard Gaussian and uncorrelated with ϵ_t. Then, the model has the following state-space representation,

$$\begin{aligned} (X_t - \bar{X}) &= \Phi_X(X_{t-1} - \bar{X}) + \Sigma_X \epsilon_t \\ P_t &= \Psi(X_t, C_t, Z_t) + \nu_t, \end{aligned} \tag{10}$$

where $X_t \equiv [F_t^T \; L_t]^T$ and Ψ is the (non-linear) mapping of cash flows C_t, bond characteristics, Z_t, and current states, X_t, into prices, P_t. The measurement errors, ν_t, have diagonal covariance matrix R.

Estimation of this system is challenging since the joint density of factors and prices is unknown. Various strategies to deal with non-linear state-space systems have been proposed in the filtering literature: the Extended Kalman Filter (EKF), the Particle Filter (PF) and more recently the Unscented Kalman Filter (UKF).[1] The UKF is described in greater detail in Appendix C. In practice, it delivers second-order accuracy with no increase in computing costs relative to the EKF. Moreover, analytical derivatives are not required. The UKF has been introduced in the term structure literature by Leippold and Wu (2003) and in the foreign exchange literature by Bakshi et al. (2005).

To set up notation, we state the standard Kalman filter algorithm as applied to our model. We then explain how the unscented approximation helps overcome the challenge posed by a non-linear state-space system. First, consider the case where Ψ is linear in X and where state variables and bond prices are jointly Gaussian. In this case, the Kalman recursion provides optimal estimates of current state variables given past and current prices. The recursion works off estimates of state variables and their associated

[1] See Julier et al. (1995), Julier and Uhlmann (1996) and Wan and der Merwe (2001) for a textbook treatment. Another popular approach bypasses filtering altogether. It assumes that some prices are observed without errors and obtains factors by inverting the pricing equation. In our context, the choice of maturities and liquidity types that are not affected by measurement errors is not innocuous and impacts estimates of the liquidity factor.

MSE from the previous step,

$$\hat{X}_{t+1|t} \equiv E\left[X_{t+1}|\Im_t\right],$$
$$Q_{t+1|t} \equiv E\left[(\hat{X}_{t+1|t} - X_{t+1})(\hat{X}_{t+1|t} - X_{t+1})^T\right], \tag{11}$$

where $\Im_t$ belongs to the natural filtration generated by bond prices. The associated predicted bond prices, and MSE, are given by

$$\hat{P}_{t+1|t} \equiv E\left[P_{t+1}|\Im_t\right] = \Psi(\hat{X}_{t+1|t}, C_{t+1}, Z_{t+1}), \tag{12}$$
$$R_{t+1|t} \equiv E\left[(\hat{P}_{t+1|t} - P_{t+1})(\hat{P}_{t+1|t} - P_{t+1})^T\right] = \Psi(\hat{X}_{t+1|t}, C_{t+1}, Z_{t+1})^T \hat{Q}_{t+1|t} \Psi(\hat{X}_{t+1|t}, C_{t+1}, Z_{t+1}) + R, \tag{13}$$

using the linearity of Ψ. The next step compares predicted bond prices to observed prices and update the state variables and their MSE,

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1}(P_{t+1} - \hat{P}_{t+1|t}), \tag{14}$$
$$Q_{t+1|t+1} = Q_{t+1|t} + K'_{t+1}(R_{t+1|t})^{-1} K_{t+1}, \tag{15}$$

where

$$K_{t+1} \equiv E\left[(\hat{X}_{t+1|t} - X_{t+1})(\hat{P}_{t+1|t} - P_{t+1})'\right] = Q_{t+1|t}\Psi(\hat{X}_{t+1|t}, C_{t+1}, Z_{t+1}), \tag{16}$$

measures co-movements between pricing and filtering errors. Finally, the transition equation gives us a conditional forecast of X_{t+2},

$$\hat{X}_{t+2|t+1} = \Phi_X \hat{X}_{t+1|t+1}, \tag{17}$$
$$Q_{t+2|t+1} = \Phi'_X Q_{t+1|t+1} \Phi_X + \Omega_X. \tag{18}$$

The recursion delivers series $\hat{P}_{t|t-1}$ and $R_{t|t-1}$ for $t = 1, \cdots, T$. Treating $\hat{X}_{1|0}$ as a parameter, and setting $R_{1|0}$ equal to the unconditional variance of measurement errors, the sample log-likelihood is

$$L(\theta) = \sum_{t=1}^{T} l(P_t; \theta) = \sum_{t=1}^{T} \left[\log \phi(\hat{P}_{t+1|t}, R_{t+1|t})\right], \tag{19}$$

where $\phi(\cdot, \cdot)$ is the multivariate Gaussian density.

However, because $\Psi(\cdot)$ is not linear, equations 12 and 13 do not correspond to the conditional expectation of prices and the associated MSE. Also, 16 does not correspond to the conditional covariance between pricing and filtering errors. Still, the updating equations 14 and 15 remain justified as optimal linear projections. Then, we can recover the Kalman recursion provided we obtain approximations of the relevant conditional moments. This is precisely what the unscented transformation achieves, using a small deterministic sample from the conditional distribution of factors while maintaining a higher order approximation than linearization. We can then use the likelihood given in 19, but in a QML context. The QML estimator is asymptotically normal and it is a strongly consistent estimator for, θ_0, the parameter vector that minimizes the Kullback-Liebler Information Criterion. This hold even if the model is mis-specified (we use an approximate filter). It may or may not be consistent for particular parameters of interest. Lund (1997) studies the properties of the QML in a similar problem as ours with a non-linear pricing function of coupon bonds. Using an iterated EKF technique he shows that the biases associated with the QML are very small. Christoffersen et al. (2007) show that the UKF improves filtering with respect to EKF.

From White (1982), we have that $\hat{\theta} \approx N(\theta_0, T^{-1}\Xi)$ where $\hat{\theta}$ is the QML estimator. The covariance matrix is

$$\Xi = E\left[\left(\zeta_H \zeta_{OP}^{-1} \zeta_H\right)^{-1}\right], \tag{20}$$

where ζ_H and ζ_{OP} are the alternative representations of the information matrix, in the Gaussian case. These can be consistently estimated via their sample counterparts. We have

$$\hat{\zeta_H} = -T^{-1}\left[\frac{\partial^2 L(\hat{\theta})}{\partial\theta\partial\theta'}\right] \tag{21}$$

and

$$\hat{\zeta_{OP}} = T^{-1}\sum_{t=1}^{T}\left[\left(\frac{\partial l(t,\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l(t,\hat{\theta})}{\partial\theta}\right)^T\right]. \tag{22}$$

Finally, the model implies some restrictions on the parameter space. In particular, ϕ_l and diagonal elements of Φ_f must lie in $(-1, 1)$ while κ and λ must remain positive. In practice, large values of κ or λ lead to numerical difficulties and are excluded. Finally, we maintain the second covariance contour of state variables inside the parameter space associated with positive interest rates. The filtering algorithm often fails outside this parameter space. None of these constraints binds around the optimum and estimates remain unchanged when the constraints are relaxed. Estimation is implemented in MATLAB via the *fmincon* routine with the medium-scale (active-set) algorithm. Different starting values were used. For standard errors computations, we obtain the final Hessian update (BFGS formula) and each observation gradient is obtained through a centered finite difference approximation evaluated at the optimum.

C Unscented Kalman Filter

The UKF is based on a method for calculating statistics of a random variable which undergoes a nonlinear transformation. It is based on an approximation to any non-linear transformation of a probability distribution. It starts with a well-chosen set of points with given sample mean and covariance. The nonlinear function is then applied to each point and moments are computed from transformed points. This approach has a Monte Carlo flavor but the sample is drawn according to a specific deterministic algorithm. It has been introduced in Julier et al. (1995) and Julier and Uhlmann (1996) (see Wan and der Merwe (2001) for textbook treatment) and was first imported in finance by Leippold and Wu (2003).

Given $\hat{X}_{t+1|t}$ a time-t forecast of state variable for period $t+1$, and its associated MSE $\hat{Q}_{t+1|t}$ the unscented filter selects a set of Sigma points in the distribution of $X_{t+1|t}$ such that

$$\bar{\mathbf{x}} = \sum_i w^{(i)} x^{(i)} = \hat{X}_{t+1|t}$$
$$\mathbf{Q}_x = \sum_i w^{(i)} (x^{(i)} - \bar{\mathbf{x}})(x^{(i)} - \bar{\mathbf{x}})' = \hat{Q}_{t+1|t}.$$

Julier et al. (1995) proposed the following set of Sigma points,

$$x^{(i)} = \begin{cases} \bar{\mathbf{x}} & i = 0 \\ \bar{\mathbf{x}} + \left(\sqrt{\frac{N_x}{1-w^{(0)}} \sum_x}\right)_{(i)} & i = 1, \dots, K \\ \bar{\mathbf{x}} - \left(\sqrt{\frac{N_x}{1-w^{(0)}} \sum_x}\right)_{(i-K)} & i = K+1, \dots, 2K \end{cases}$$

with weights

$$w^{(i)} = \begin{cases} w^{(0)} & i = 0 \\ \frac{1-w^{(0)}}{2K} & i = 1, \dots, K \\ \frac{1-w^{(0)}}{2K} & i = K+1, \dots, 2K \end{cases}$$

where $\left(\sqrt{\frac{N_x}{1-w^{(0)}} \sum_x}\right)_{(i)}$ is the i-th row or column of the matrix square root. Julier and Uhlmann (1996) use a Taylor expansion to evaluate the approximation's accuracy. The expansion of $y = g(x)$ around $\bar{x}$ is

$$\begin{aligned} \bar{y} &= E\left[g(\bar{x} + \Delta x)\right] \\ &= g(\bar{x}) + E\left[D_{\Delta x}(g) + \frac{D^2_{\Delta x}(g)}{2!} + \frac{D^3_{\Delta x}(g)}{3!} + \cdots\right] \end{aligned}$$

where the $D^i_{\Delta x}(g)$ operator evaluates the total differential of $g(\cdot)$ when perturbed by Δx, and evaluated at $\bar{x}$. A useful representation of this operator in our context is

$$\frac{D^i_{\Delta x}(g)}{i!} = \frac{1}{i!}\left(\sum_{j=1}^{n} \Delta x_j \frac{\partial}{\partial x_j}\right)^i g(x)\bigg|_{x=\bar{x}}$$

Different approximation strategies for $\bar{y}$ will differ by either the number of terms used in the expansion or the set of perturbations Δx. If the distribution of Δx is symmetric, all odd-ordered terms are zero. Moreover, we can re-write the second terms as a function of the covariance matrix P_{xx} of Δx,

$$\bar{y} = g(\bar{x}) + \left(\nabla^T P_{xx} \nabla\right) g(\bar{x}) + E\left[\frac{D^4_{\Delta x}(g)}{4!} + \cdots\right]$$

Linearisation leads to the approximation $\hat{\bar{y}}_{lin} = g(\bar{x})$ while the unscented approximation is exact up to the third-order term and the σ-points have the correct covariance matrix by construction. In the Gaussian case, Julier and Ulhmann Julier and Uhlmann (1996) show that same-variable fourth moments agree as well and that all other moments are lower than the true moments of Δx. Then, approximation errors of higher order terms are necessarily smaller for the UKF than for the EKF. Using a similar argument, but for approximation of the MSE, Julier and Uhlmann (1996) show that linearisation and the unscented transformation agree with the Taylor expansion up to the second-order term and that approximation errors in higher-order terms are smaller for the UKF.

D Data

We use end-of-month prices of U.S. Treasury securities from the CRSP data set. We exclude callable bonds, flower bonds and other bonds with tax privileges, issues with no publicly outstanding securities, bonds and bills with less than 2 months to maturity, issues that were 5 years old or more, and observations with either bid or ask prices missing. Our sample covers the period from January 1986 to December 2009. In the construction of pairs at each maturity, we excluded suspicious quotes that show up as outliers relative to surrounding quotes. These are given in Table 1. We also exclude CRSP ID #20040304.400000 since its maturity date precedes its issuance date, as dated by the U.S. Treasury. Including any of these observations do not affect our results significantly.

Table 1: **Outliers**

The following observations were excluded in the process of constructing pairs of bond prices. These are most likely clerical errors since they appear as outliers relative to surrounding quotes on the same observation day.

CRSP ID	Date
#19920815.107250	August 31^{st} 1987
#19950331.203870	December 30^{th} 1994
#19980528.400000	May 30^{th} 1998
#20011130.205870	October 31^{th} 1997
#20030228.205500	February 26^{th} 1999
#20041031.202120	November 29^{th} 2002
#20070731.203870	May 31^{st} 2006
#20080531.204870	November 30^{th} 2007

A few more observations were excluded when the most recent issue was not special, contradicting the maintained assumption that bond prices do not increase with age. Before 2008, the only exclusion is CRSP ID #20130815.204250 since it is never special. In 2008, the most recent 10-year issue is typically *not* the most expensive in its maturity group. Table 2 details what bonds were selected for that maturity group in 2008. Note, however, that in each case, the relationship between bond prices and age remained stable for the other bonds. Is it only the relationship between the few most recent issues that appears to be varying.

E Parameter Estimates

We impose that $K^{\mathbb{P}}$ is diagonal but the presence of the off-diagonal elements does not affect our results. Moreover, CDR show that allowing for an unrestricted matrix $K^{\mathbb{P}}$ deteriorates out-of-sample performance. Σ_f is lower triangular for identification purposes. Table 3(a) and Table 3(b) report parameter estimates. For standard errors, we reports two figures, a robust one using both the Hessian covariance matrix and the outerproduct of the scores, which we call QML, and a second one based on the outerproduct of the

Table 2: **Bond Pairs for the 10-year Maturity Group in 2008**

CRSP ID number of bonds forming a pair in the 10-year maturity group from January to December 2008.

Date	CRSP ID#	
	New	Old
01/2008	20171115.204250	20170815.204750
02/2008	20180215.203500	20171115.204250
03/2008	20180215.203500	20171115.204250
04/2008	20180215.203500	20171115.204250
05/2008	20180515.203875	20171115.204250
06/2008	20180215.203500	20171115.204250
07/2008	20180215.203500	20171115.204250
08/2008	20180215.203500	20170815.204750
10/2008	20180215.203500	20170815.204750
11/2008	20180515.203875	20170815.204750
12/2008	20180515.203875	20180215.203500

scores only, which we call OP. The first measure probably overestimates the variability[2], while the second one surely underestimates it. Therefore we decided to report both metrics.

In the benchmark model, the estimate of the curvature parameter is $\hat{\lambda} = 0.6786$, when time periods are measured in years. This estimate pins the maximum curvature loading at a maturity close to 30 months. Its QML and OP standard errors are 0.0305 and 0.0044, respectively. The results imply average short and long term discount rates of 3.73% and 5.45%, respectively. Standard deviations of pricing errors are given by

$$\begin{aligned} \sigma_r(M_n) = \quad & \underset{(0.017,\ 0.0012)}{0.0229} \quad + \quad \underset{(0.021,\ 0.0006)}{0.0284} \times M_n, \end{aligned}$$

with QML and OP standard errors for each parameter. This implies standard deviations of \$0.05 and \$0.31 for maturities of 1 and 10 years, respectively. Using durations of 1 and 7 years, this translates into yield errors of 5.1 and 4.4 bps.

Estimate for the curvature parameter in the model with liquidity is $\hat{\lambda} = 0.7304$ with QML and OP standard errors of 0.0857 and 0.0043. The standard deviations of measurement errors are given by

$$\sigma_r(M_n) = \underset{(0.016, 0.001)}{0.0227+} \quad \underset{(0.0021, 0.0006)}{0.0251} \times M_n,$$

with QML and OP standard errors for each parameter in parenthesis. Then, standard deviations are \$0.048 and \$0.274 for bonds with one and ten years to maturity, respectively. Using durations of 1 and 7, this translates into standard deviations of 4.8 and 3.9 bps when measured in yields compared with 5.1 bps an 4.4 bps from the benchmark results. Overall, parameter estimates and latent factors are relatively unchanged compared to the benchmark model.

[2]The Hessian is not available in closed-form and a numerical approximation for the second derivative of the entire likelihood introduces errors.

Table 3: **Parameter Estimates - Transition Equations**

Panel (a) presents estimation results for the AFENS model without liquidity. Panel (b) presents estimation results for the AFENS model with liquidity. For each parameter, the first standard error (in parentheses) is computed from the QMLE covariance matrix while the second is computed from the outer product of scores. End-of-month data from CRSP (1985:12-2007:12).

(a) Benchmark Model

	$\bar{F}$	K		Σ ($\times 10^2$)	
	0.0545	0.169	0.68		
Level	(0.0136)	(0.177)	(0.42)		
	(0.0093)	(0.069)	(0.03)		
	-0.0172	0.182	0.76	0.84	
Slope	(0.0277)	(0.088)	(0.75)	(0.46)	
	(0.013)	(0.071)	(0.06)	(0.04)	
	-0.0128	0.891	-0.14	0.41	2.31
Curvature	(0.0061)	(0.860)	(1.86)	(1.64)	(0.66)
	(0.0045)	(0.283)	(0.15)	(0.17)	(0.13)

(b) Model with Liquidity

	$\bar{F}$	K		Σ ($\times 10^2$)	
	0.0576	0.198	0.85		
Level	(0.0165)	(0.165)	(0.86)		
	(0.0154)	(0.098)	(0.02)		
	-0.0167	0.222	-0.81	0.85	
Slope	(0.0092)	(0.293)	(0.85)	(0.44)	
	(0.0165)	(0.145)	(0.06)	(0.05)	
	-0.0189	0.887	0.57	0.25	2.27
Curvature	(0.0057)	(1.414)	(0.82)	(1.91)	(1.66)
	(0.0088)	(0.325)	(0.13)	(0.20)	(0.12)

	$\bar{L}$	ϕ_l	σ_l
	0.32	0.955	0.06
Liquidity	(0.42)	(0.034)	(0.066)
	(0.09)	(0.021)	(0.011)

F Response of Treasury Yields to Funding Liquidity Shocks

The value of funding liquidity predicts future excess bond returns conditional on the level, slope and curvature. This raises important questions at the heart of a recent literature on affine term structure models (e.g. Joslin et al. (2010) and Duffee (2011)). In general, we expect increases (decreases) in expected excess bond returns to be associated with decreases (increases) in current bond prices to generate potential for higher returns. This is the same mechanism than the Campbell-Shiller decomposition of stock returns into news about future cash flows, about futures expected excess returns and about future short rates. (Note that changes in expected cash flows play no role when we focus on nominal bond returns.) However, in the context of Gaussian dynamic term structure models, and in the absence of pricing or measurement errors in yields, the predictive content of any state variable is captured if we include a sufficient number of yields (or yield factors) in predictive regressions. Observing a sufficient number of linear combinations of prices reveals all that we need to know about expected excess bond returns.

One exception is the knife-edge case where the higher (lower) term premium is exactly offset by a lower (higher) future path for the short rate. In this case, some risk factors may affect the risk premium but not the cross-section of yields. Could it be that L_t is unspanned but affects the price of F_t risk? This possibility relies on a number of maintained hypothesis about the number of factors, their dynamics, the absence of friction and linearity. In fact, Duffee (2011) warns not to take the knife-edge restrictions literally. He argues that some factor may be hard to measure due to distortions on the bond markets. In particular, he points at sources of small, transitory and idiosyncratic noises.[3] In contrast, we introduce small, *persistent* and *common* deviations between old and new bonds. These deviations are computed relative to an idealized yield curve implied by a frictionless no-arbitrage model. They are hidden from models assuming transitory and uncorrelated errors and where estimation is based on zero-coupon curves.

Can we reconcile this evidence that L_t predicts risk premium with the underlying frictionless zero-coupon curve that we estimated? Appendix G introduces a generalized no-arbitrage Nelson-Siegel term structure model in discrete-time where the state vector is extended to include additional variables. These variables can have arbitrary loadings and their historical dynamics is unrestricted. It shows that the predictability from L_t cannot be due to its effect on the prices of F_t risk whenever F_t and L_t shocks are uncorrelated as we assumed. Allowing for a general correlation and auto-correlation structure opens interesting cases. L_t may or may not be priced. It may enter the prices of F_t risks. It may play a role in the dynamics of F_t. It may be a component of the short rate equation. This points toward a detailed analysis of L_t's role for Treasury bonds. This is clearly beyond the scope of the paper and we leave these important questions for future research. Importantly, Section H below shows that we expect the filtered factors $\hat{F}_t$ to capture any contemporaneous effect of L_t on yields whenever we estimate a restricted model where L_t does not affect yields. This is what latent factor models do after all. This arises if the patterns of the loadings of yields on L_t across maturities can be obtained (approximately) as a linear combination

[3]See Duffee (2011), Section 2.7. "First, there are market imperfections that distort bond prices, such as bid/ask spreads. Second, there are market imperfections that distort payoffs to bonds (and thus distort what investors will pay for bonds), such as special RP rates. Third, there are distortions created by the mechanical interpolation of zero-coupon bond prices from coupon bond prices."

of level, slope and curvature loadings.[4]

In any case, the evidence does not preclude that funding liquidity is spanned by yields. First, PCAs computed from yields span a substantial share of the filtered liquidity factor variations. Principal Components from the Fama-Bliss CRSP files (T-bills rates at maturities of 1, 2, 3, 4, 5 and 6 months. Zero-coupon yields at maturities of 1, 2, 3, 4 and 5 years) span 47% of funding liquidity variations. PCAs from the dataset of Gurkaynak et al. (2006) (maturities of 3, 6, 9 and 12 months as well as 1, 2, 3, 4, 5, 7, 10 years) span 54% of the funding liquidity variations. Imperfect spanning ($R^2 < 1$) may be consistent with measurement errors in yields and filtered estimates of L_t.

Second, although we cannot clean the filtered factors $\hat{F}_t$ from the omitted effect of L_t, we can ask what part of $\hat{F}_t$ innovations comes from its innovations in L_t. We use regressions to check the contemporaneous relationship between bond price changes (i.e. bond returns) computed as $r^m_{t+1} = \log(D^{(m-1)}_{t+1}/D^{(m)}_t)$ and innovations in state variables, $\epsilon_{t+1} = X_{t+1} - E_t[X_{t+1}]$. Table 4 provides results from contemporaneous regressions of monthly bond returns on funding liquidity and term structure shocks. Regressors are normalized for ease of interpretation. We also include lagged state variables, L_{t-1} and F_{t-1}, as conditioning information. The evidence is unambiguous. An unexpected shock to the value of funding liquidity is associated with positive bond returns contemporaneously. For a 10-yr bond, a shock to funding liquidity is associated with a 0.3% monthly returns. One interpretation is that Treasury bonds provide a hedge against funding liquidity shocks. Consistent with predictability regressions, this higher current value of funding liquidity is associated with lower expected returns in the future. This holds for every maturity we considered. Coefficients of lagged term structure factors and of their innovations have the expected sign.[5] Note that inclusion of lagged factors does not affect point estimates for innovation coefficients since these two sets of regressors are uncorrelated by construction. Instead, this makes inference more reliable since a short-window kernel can then be used for the purpose of computing Newey-West standard errors.

The results are based on a joint VAR dynamics for the term structure and liquidity factors. We identify X_t shocks via the Cholesky decomposition of their covariance matrix where L_t is ordered before the term structure factors.[6] This decomposition let the data speaks about the correlation between innovations. In contrast, any other ordering precludes the effect of innovations to $\hat{L}_t$ on some or all aspects of yield changes *by assumption*. There is no reason, a priori, that shocks to the funding liquidity factor do not affect term structure factors. Whether it does is an empirical question. The VAR dynamics was re-estimated from

[4] If this is not the case, the effect of L_t (or any other missing factor) on yields is likely to be small since yield variations can be accurately described in term of level, slope and curvature changes.

[5] Coefficient estimates imply that Level shocks are associated with a -2.57% returns for a 10-year discount bond. Slope and Curvature shocks are associated with -0.39% and -0.68% returns, respectively. Coefficient estimates on lagged term structure factors are positive and significant (i.e. 0.57, 0.07% and 0.18%). This is consistent with the observation that low returns today imply higher expected future returns. Unsurprisingly, term structure factors explain a large proportion of yield variations since the estimates maximize the likelihood of observed yields. Their t-statistics in yield changes regressions are necessarily large and are not reported to preserve space.

[6] The paper considers a restricted specification where L_t does not enter the short rate directly, does not affect the dynamics of F_t and its shocks are uncorrelated to F_t shocks. We took this approach because it identified the latent L_t solely to price deviations in the cross-section of bond ages that cannot be linked to cash flow differences. This eases the interpretation of the evidence as there is no channel through which the filtered value of L_t mix information from the term structure with information from the cross-section of age. Leaving these channels open would have raised doubts about the source of its information content for future excess bond returns.

filtered factors but with an unrestricted auto-regressive matrix. Using estimates of the restricted VAR in the paper only makes the evidence stronger since the correlation structure between elements of X_t is entirely attributed to correlation between innovations. Results are based on returns from the GSW dataset. The evidence is stronger if we compute returns using discount bond prices from the model. For completeness, we include innovations to all state variable in the regressions. Finally, very high R^2s are not surprising since the latent system dynamics was obtained by maximizing the conditional log-likelihood of each monthly yield changes.

G Generalized Nelson-Siegel Representations

G.1 Adding State Variables

We show that the argument in CDR extends easily to the case where the state vector include the Nelson-Siegel factors, F_t, as well as other factors, Z_t, entering the equation for yields.[7] This is most easily seen in a discrete-time version of the model where the vector, $X_t = [F_t' \; Z_t']'$, follows a Gaussian VAR(1) under $\mathbb{Q}$,

$$X_{t+1} = \mu_x^{\mathbb{Q}} + \Phi_x^{\mathbb{Q}} X_t + \epsilon_{x,t}^{\mathbb{Q}},$$

with $\epsilon_{f,t}^{\mathbb{Q}} \sim^{\mathbb{Q}} N(0, \Omega)$, where

$$\mu_x^{\mathbb{Q}} = \begin{bmatrix} \mu_f \\ \mu_z \end{bmatrix} \quad \Phi_x^{\mathbb{Q}} = \begin{bmatrix} \Phi_f & \Phi_{f,z} \\ \Phi_{z,f} & \Phi_z \end{bmatrix} \quad \Omega_x = \begin{bmatrix} \Omega_f & \Omega_{f,z} \\ \Omega_{z,f} & \Omega_z \end{bmatrix},$$

and, therefore, the conditional Laplace transform is

$$E_t^{\mathbb{Q}} \left[\exp\left\{ u_f' F_{t+1} + u_z' Z_{t+1} \right\} \right] = \exp\left\{ \frac{u_x' \Omega u_x}{2} + u_x' (\mu_x^{\mathbb{Q}} + \Phi_x^{\mathbb{Q}} X_t) \right\}.$$

Suppose that the short rate is given by

$$r_t = y_t^{(1)} = \delta_f' F_t + \delta_z' Z_t,$$

then the price of a zero-coupon bond price with maturity $m = 1, 2, \ldots$ is

$$P_t^{(m)} = \exp\left\{ A(m) + B_f(m)' F_t + B_z(m)' Z_t \right\}$$

with coefficients given by the following recursions

$$A(m+1) = A(m) + B(m)' \mu_x^{\mathbb{Q}} + \frac{B_x(m)' \Omega B_x(m)}{2} \tag{23}$$

$$B_f(m+1) = \Phi_f^{\mathbb{Q}\prime} B_f(m) + \Phi_{z,f}^{\mathbb{Q}\prime} B_z(m) - \delta_f \tag{24}$$

$$B_z(m+1) = \Phi_{f,z}^{\mathbb{Q}\prime} B_f(m) + \Phi_z^{\mathbb{Q}\prime} B_z(m) - \delta_z \tag{25}$$

[7] The discussion in this Section follows from comments and suggestions from Bruno Feunou.

Table 4: **Funding Liquidity and Contemporaneous Treasury Bond Returns**

Results from the contemporaneous regressions,

$$r_{t+1}^{(m)} = \alpha + \beta_u \top u_{t+1} + \beta_X \top X_t + \epsilon_{t+1},$$

where $r_{t+1}^{(m)}$ is the monthly log-returns from an off-the-run zero-coupon bond with maturity m, where $X_t = [L_t \; F_t]$ stacks the funding liquidity and term structure factors from the AFENS model with liquidity, and where u_{t+1} is the Cholesky decomposition of $X_{t+1} - E_t[X_{t+1}]$. Regressors are demeaned and divided by their standard deviations. Newey-West standard errors (3 lags) in parentheses. End-of-month data from CRSP (1985:12-2007:12).

	Bond Maturity									
Variable	3m	6m	9m	1y	2y	3y	4y	5y	7y	10y
Cst	0.41	0.43	0.45	0.46	0.52	0.57	0.61	0.65	0.71	0.78
	Funding Liquidity Innovations									
$u_{liq,t+1}$	0.011	0.013	0.008	0.004	0.017	0.057	0.103	0.145	0.216	0.299
	(5.746)	(7.319)	(3.932)	(1.749)	(4.647)	(10.530)	(11.317)	(12.525)	(15.569)	(6.688)
	Nelson-Siegel Factor Innovations									
$u_{lvl,t+1}$	-0.006	-0.031	-0.070	-0.119	-0.370	-0.658	-0.953	-1.246	-1.806	-2.571
$u_{slp,t+1}$	-0.028	-0.076	-0.123	-0.164	-0.278	-0.337	-0.368	-0.384	-0.395	-0.395
$u_{crv,t+1}$	-0.003	-0.031	-0.073	-0.123	-0.336	-0.520	-0.653	-0.736	-0.785	-0.680
	Funding Liquidity									
L_t	-0.018	-0.025	-0.038	-0.054	-0.131	-0.204	-0.266	-0.316	-0.392	-0.467
	(-10.091)	(-14.852)	(-21.013)	(-25.969)	(-45.261)	(-44.930)	(-37.740)	(-34.771)	(-30.051)	(-13.147)
	Nelson-Siegel Factors									
$F_{1,t}$	0.143	0.160	0.175	0.190	0.250	0.306	0.355	0.398	0.476	0.584
$F_{2,t}$	0.146	0.148	0.147	0.146	0.143	0.137	0.127	0.116	0.094	0.074
$F_{3,t}$	0.003	0.004	0.006	0.010	0.033	0.060	0.087	0.111	0.149	0.184
R^2	97.7	98.3	98.4	98.6	99.3	99.4	99.4	99.4	99.2	97.6

with $A(0) = B_f(0) = B_z(0) = 0$. Suppose that $\Phi^{\mathbb{Q}}_{z,f} = 0$. Then, if δ_f and $\Phi^{\mathbb{Q}}_f$ are restricted to

$$\delta_f = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda}}{\lambda} \\ \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \end{bmatrix}' \quad \Phi^{\mathbb{Q}}_f = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda} & \lambda e^{-\lambda} \\ 0 & 0 & e^{-\lambda} \end{pmatrix}, \tag{26}$$

respectively, then $b(m) = -B(m)/m$ corresponds to Nelson-Siegel loadings as in Equation 4 but for integer $m = 1, 2, 3, \ldots$. Moreover, we recover the same constant as in Equation 6 (i.e. $\tilde{a}(m)$) by imposing $\Omega_{z,f} = \Omega_{f,z} = 0$ so that innovations to F_t and Z_t are uncorrelated (and we still assume $\mu^{\mathbb{Q}} = 0$ for identification purposes). This corresponds to the model estimated in the paper.

G.2 Change of Measure

Assume, further, that the state vector also follows a Gaussian VAR(1) dynamics under $\mathbb{P}$ with parameters μ, Φ_x and Ω. This is equivalent to choosing a log-normal pricing kernel,

$$M_{t+1} = \exp\left\{-r_t - \frac{1}{2}\eta'\Omega\eta - \eta'\left(\epsilon_{t+1} - \eta_t\right)\right\},$$

with linear prices of risk, η_t,

$$\eta_t = \Sigma\left(\eta_0 + \eta X_t\right),$$

where $\Omega = \Sigma\Sigma'$. Prices of risk parameters are given by

$$\begin{aligned} \eta_0 =& \Sigma^{-1}\left(\mu - \mu^{\mathbb{Q}}\right) \\ \eta =& \Sigma^{-1}\left(\Phi_x - \Phi^{\mathbb{Q}}_x\right). \end{aligned} \tag{27}$$

While the dynamics under the historical measure is left unrestricted, as in CDR, the dynamics under $\mathbb{Q}$ is restricted (i.e. $\Phi^{\mathbb{Q}}_{z,f} = 0$ and $\Phi^{\mathbb{Q}}_f$ is given by Equation 26). Then, for any configuration of parameters μ_x, Φ_x and Ω, Equation 27 shows what price of risk parameters are consistent with a yield equation where loadings on F_t are given by the Nelson-Siegel representation. In particular, the restrictions $\Omega_{z,f} = \Omega_{f,z} = 0$ imply that $\eta_{f,z}$ and $\eta_{z,f}$ are redundant since shocks to Z_t are not sources of risk in this case.

G.3 Span Restrictions for Z_t

The general case allows Z_t to enter the equation for yields. Following Duffee (2011), we can easily derive conditions under which Z_t is not spanned by the yield curve and remains hidden. We maintain the assumptions made above to ensure that F_t enter the yield curve with NS loading. Looking back at the recursions in Equation 23, it is easy to see that Z_t does not enter the yield curve (i.e. $B_z(m) = 0$ for any $m \geq 0$) if and only if $\delta_z = 0$ and $\Phi^{\mathbb{Q}}_{f,z} = 0$. In particular, the restrictions can be satisfied irrespective of the historical dynamics for an appropriate choice of the price of risk parameters.

$$\begin{bmatrix} \eta_f & \eta_{f,z} \\ \eta_{z,f} & \eta_z \end{bmatrix} = \begin{bmatrix} \Sigma_f & \Sigma_{f,z} \\ \Sigma_{z,f} & \Sigma_z \end{bmatrix}^{-1} \left(\begin{bmatrix} \Phi^{\mathbb{Q}}_f & 0 \\ 0 & \Phi^{\mathbb{Q}}_z \end{bmatrix} - \begin{bmatrix} \Phi_f & \Phi_{f,z} \\ \Phi_{z,f} & \Phi_z \end{bmatrix} \right), \tag{28}$$

Next, consider the case estimated in the paper. We have that $Z_t = l_t$ is univariate and,

$$\eta_f = \Sigma_f^{-1}\left(\Phi_f^{\mathbb{Q}} - \Phi_f\right) \qquad \eta_l = \sigma_l^{-1}\left(\Phi_l^{\mathbb{Q}} - \phi_l\right). \tag{29}$$

H Kalman Filtering and Missing Factors

Relative to the general framework abvove, the paper imposes the following restrictions,

$$\Phi_{z,f}^{\mathbb{Q}} = \Phi_{z,f}^{\mathbb{Q}} = 0$$

and

$$\delta_z = 0 \qquad \Phi_{z,f}^{\mathbb{P}} = \Phi_{z,f}^{\mathbb{P}} = 0$$

on the $\mathbb{Q}$-dynamics and the $\mathbb{P}$ dynamics, respectively. It also imposes that $\Omega_{f,z} = 0$. We treat these restrictions in turns.

H.1 Effect of $\Omega_{f,z} = 0$

This restriction is harmless and can be seen as a rotation of the innovations. The required rotation from an arbitrary unrestricted matrix Ω_x is not unique. In particular, we can choose a rotation that leaves either ϵ_{t+1}^f or ϵ_{t+1}^z unchanged. Define the matrix $\bar{A}$,

$$\bar{A} = \begin{bmatrix} I_{n_f} & \Omega_{f,z}\Omega_z^{-1} \\ 0 & I_{n_z} \end{bmatrix}.$$

Then, the transformed innovations, $\bar{A}^{-1}\epsilon_{t+1}$, have a block diagonal covariance matrix given by

$$\bar{A}^{-1}\Omega_x(\bar{A}^{-1})' = \begin{bmatrix} \Omega_f - \Omega_{f,z}\Omega_z^{-1} & 0 \\ 0 & \Omega_z \end{bmatrix}.$$

This particular rotation leaves the innovation of Z_t unchanged but re-define the innovations to the term structure factors as the residuals from a projection.

H.2 Effect of δ_z, $\Phi_{f,z}^{\mathbb{P}} = 0$ *and* $\Phi_{f,z}^{\mathbb{Q}} = 0$

This restrictions affects the Kalman filter substantially. With no loss of generality, we consider a filter for the centered state vector $X_t - \bar{X}$,

$$(X_{t+1} - \bar{X}) = \Phi_x(X_t - \bar{X}) + \epsilon_{t+1},$$

where, as above, ϵ_{t+1} has a block diagonal covariance matrix, Ω. The initial values for the filter are

$$\hat{F}_{0|0} = E[F_t] = \bar{F} \quad \text{and} \quad Q_{0|0}^f = \Lambda_f, \tag{30}$$

where $vec(\Lambda_f) = (I - \Phi_f \otimes \Phi_f')^{-1} vec(\Omega_f)$ is the (true) unconditional variance of F_t. The Kalman forecast is given by

$$\hat{F}_{1|0} = \bar{F} + \Phi_f(\hat{F}_{0|0} - \bar{F}),$$

although the unrestricted forecast should account for the lagged effect of Z_t. The associated Mean Squared Error is,

$$Q^f_{1|0} = E\left[\left(F_1 - \hat{F}_{1|0}\right)\left(F_1 - \hat{F}_{1|0}\right)'\right] = [\Phi_f \ \Phi_{f,z}]\, Q^x_{0|0}\, [\Phi_f \ \Phi_{f,z}]' + \Omega_f,$$

where $Q^x_{0|0} = \Lambda$ is the long-run covariance matrix of X_t. Parameters associated with Z_t play a role in the MSE for $\hat{F}_{t+1|t}$ because of the omitted term in the forecast step. Next, the predicted yields in the measurement equations are given by

$$\hat{Y}_{1|0} = A(\lambda, \Omega_f) + B_f(\lambda)'\hat{F}_{1|0},$$

although the unrestricted prices should account for the effect of Z_{t+1}. It follows from standard results that the optimal linear update, $\hat{F}_{1|1}$, is given by,

$$\hat{F}_{1|1} = \hat{F}_{1|0} + E\left[\left(F_1 - \hat{F}_{1|0}\right)\left(Y_1 - \hat{Y}_{1|0}\right)'\right] E\left[\left(Y_1 - \hat{Y}_{1|0}\right)\left(Y_1 - \hat{Y}_{1|0}\right)'\right]^{-1} \times \left(Y_1 - \hat{Y}_{1|0}\right). \tag{31}$$

where we can evaluate each term in turns. The measurement errors are,

$$Y_1 - \hat{Y}_{1|0} = B_f(\lambda)'(F_1 - \hat{F}_{1|0}) + B_z(\Theta)'Z_1 + \nu_1,$$

which include the usual measurement errors, ν_1, and the state forecast errors, $(F_1 - \hat{F}_{1|0})$, as well as a new term omitted from the yield equation, $B_z(\Theta)'Z_1$. This term arises because of the restrictions that Z_t is unspanned (i.e. $\Phi^{\mathbb{Q}}_{f,z} = 0$ and $\delta_z = 0$). The covariance term is given by,

$$\begin{aligned} E\left[\left(F_1 - \hat{F}_{1|0}\right)\left(Y_1 - \hat{Y}_{1|0}\right)'\right] &= [\Phi_f \ \Phi_{f,z}]\, Q^x_{0|0}\Phi_x' B_x(\theta) + \Omega_f B_f(\lambda) \\ &= S_f \Phi_x Q^x_{0|0} \Phi_x' B_x(\theta) + S_f \Omega_x B_x(\theta) \end{aligned} \tag{32}$$

where $S_f = [I_f \ 0_{f,z}]$ is a selection matrix such that $S_f \Phi_x = [\Phi_f \ \Phi_{f,z}]$ and $S_f \Omega_x B_x(\theta) = \Omega_f B_f(\lambda)$ (using the fact that Ω_x is block diagonal). The variance term is given by,

$$\begin{aligned} R^y_{1|0} &= E\left[\left(Y_1 - \hat{Y}_{1|0}\right)\left(Y_1 - \hat{Y}_{1|0}\right)'\right] \\ &= B_x(\theta)' \left[\Phi_x Q_{0|0^x} \Phi_x' + \Omega_x\right] B_x(\theta) + R, \end{aligned} \tag{33}$$

where, again, $R = var(\nu_1)$ is the measurement error variance.

H.3 What Can Latent Factor Pick Up

As implemented, the Kalman filter of $\hat{F}_{t|t}$ captures some of the effect of Z_t on yields and we have that $\tilde{F}_{t|t} \neq E[F_t|Y_{1:t}]$. But this is what latent factors should do. We expect filtered values to capture most of that effect and that $\tilde{F}_{t|t}$ is a mixture of F_t and Z_t. Recall that almost all yield variations are interpreted in terms of level, slope and curvature factors, respectively. This interpretation comes from the patterns of yield loadings on each factor across different maturities. This suggests that the loadings of actual economic variables on yields can be described as a linear combination of level, slope and curvature loadings. Formally, this assumption requires that the column of $B_z(\theta)$ are (approximately) spanned by the columns of $B_f(\lambda)$ and we have,

$$B_z(\theta) = C_{z,f} B_f(\lambda), \tag{34}$$

where $C_{z,f}$ is a $N_z \times N_f$ matrix. Note that this can only be the case if $N_z \leq N_f$ or, else, if $rank(B_z(\theta)) < N_z$. This is not an issue since we are primarily concerned with the case $N_z = 1$. In the simplest cases, inflation is often associated with level loadings and output gap with slope loadings. Here, we allow for the loadings on liquidity to combine level, slope and curvature effects via the $C_{z,f}$ matrix.

Consider again the covariance term in Equation 32. We can write,

$$E\left[\left(F_1 - \hat{F}_{1|0}\right)\left(Y_1 - \hat{Y}_{1|0}\right)'\right] = \left(Q^f_{1|0} + Q^{f,z}_{1|0} C_{z,f}\right) B_f(\lambda), \tag{35}$$

and the optimal covariance is available to the actual Kalman filter implementation by taking,

$$\tilde{Q}^f_{1|0} = Q^f_{1|0} + Q^{f,z}_{1|0} C_{z,f} = \Phi_f \tilde{Q}^f_{0|0} \Phi_f' + \tilde{\Omega}_f, \tag{36}$$

and this has tight implication for the choice of $\tilde{\Omega}_f$. From the forecast rule for $Q^x_{1|0}$, we get that

$$\tilde{\Omega}_f = \Omega_f + \Phi_{f,z} Q^{z,f}_{0|0} \Phi_f' + \Phi_f Q^{f,z}_{0|0} \Phi_{f,z} + \Phi_{f,z} Q^z_{0|0} \Phi_{f,z} + [\Phi_f \Phi_{z,f}]\, Q^x_{0|0}\, [\Phi_f \Phi_{z,f}]'\, C_{z,f}, \tag{37}$$

which shows that $\tilde{\Omega}_f$ combines the true covariance matrix of F_t with terms arising from omitting the effect of Z_t in (i) the forecast of $\hat{F}_{1|0}$ and (ii) in the yield equation. For consistency, we have that

$$vec(\tilde{Q}^f_{0|0}) = \left(I - \Phi_f \otimes \Phi_f'\right)^{-1} vec(\tilde{\Omega}_f).$$

Clearly, omitting Z_t biases parameter estimates. Equation 36 shows that one biased estimate of Ω_f delivers the optimal covariance matrix for the update step. The remaining question is whether the resulting (biased) filter captures the missing factor in yields. The update stage is given by,

$$\hat{F}_{1|1} = \hat{F}_{1|0} + \tilde{Q}^f_{1|0} B_F(\lambda) \left[B_F(\lambda)' \tilde{Q}^f_{1|0} B_F(\lambda) + R\right]^{-1} \times B_f(\lambda)'(F_1 - \hat{F}_{1|0}) + B_f(\lambda)' C_{z,f}' Z_1 + \nu_1, \tag{38}$$

where R is an order of magnitude lower than $B_F(\lambda)' \tilde{Q}^f_{1|0} B_F(\lambda)$ in typical term structure applications.

Then, approximately when R is close to zero, we have that,

$$\hat{F}_{1|1} \approx \hat{F}_{1|0} + (F_1 - \hat{F}_{1|0}) + C'_{z,k} Z_1 = F_1 + C'_{z,k} Z_1,$$

and, therefore, innovations from $\hat{F}_{t|t}$ to $\hat{F}_{t+1|t+1}$ combines innovations in the underlying F_{t+1} and Z_{t+1}.

References

Bakshi, G., P. Carr, and L. Wu (2005). Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies. *Journal of Financial Economics 87*.

Christensen, J., F. Diebold, and G. Rudebusch (2007). The affine arbitrage-free class of Nelson-Siegel term structure models. NBER W13611.

Christoffersen, P., K. Jacobs, L. Karoui, and K. Mimouni (2007). Estimating term structure modell using swap rates. McGill University.

Duffee, G. (2011). Information in (and not in) the term structure. *Review of Financial Studies*. *forthcoming.*

Gurkaynak, R., B. Sack, and J. Wright (2006). The U.S. Treasury curve: 1961 to present. Technical Report 2006-28, Federal Reserve Board.

Joslin, S., K. H. Singleton, and H. Zhu (2010). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*. forthcoming.

Julier, S. and J. Uhlmann (1996). A general method for approximating nonlinear transformations of probability distributions. Robotics Research Group, University of Oxford.

Julier, S., J. Uhlmann, and H. Durrant-Whyte (1995). A new approach for filtering nonlinear systems. *The Proceedings of the American Control Conference 3*, 1628–1632. Seattle.

Leippold, M. and L. Wu (2003). Design and estimation of multi-currency quadratic models. *European Finance Review 7*, 47–73.

Lund, J. (1997). Non-linear kalman filtering techniques for term structure models. Technical report, Aarhus School of Business.

Wan, E. and R. V. der Merwe (2001). *The unscented Kalman filter*, Chapter 7. Wiley and Sons. Kalman filtering and neural network.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), pp. 1–25.