DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

The incremental information in the yield curve about future interest rate risk

Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev

CREATES Research Paper 2021-11

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

The incremental information in the yield curve about future interest rate risk*

Bent Jesper Christensen† Mads Markvart Kjær‡ Bezirgen Veliyev§

*We are grateful to Anders B. Trolle and participants at the Conference on Computational and Financial Econometrics (CFE) in London, 2019, and the Joint Econometrics-Finance Seminar in Aarhus for useful comments. Declarations of interest: None.

†Corresponding author at: CREATES and the Dale T. Mortensen Center, Department of Economics and Business Economics, Aarhus University, and the Danish Finance Institute, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Email: bjchristensen@econ.au.dk.

‡CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Email: mads.markvart@econ.au.dk.

§ CREATES, Department of Economics and Business Economics, Aarhus University, and the Danish Finance Institute, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Email: bveliyev@econ.au.dk.

The incremental information in the yield curve about future interest rate risk

Abstract

Using high-frequency intraday futures prices to measure yield volatility at selected maturities, we find that daily yield curves carry incremental information about future interest rate risk at the long end, relative to that contained in the time series of historical volatilities. Some of the information in the yield curves is not captured by standard affine models. At the short end, time series based forecasts outperform yield curve based forecasts. Both provide utility to a risk averse investor in longer-term instruments, not in short, relative to a random walk. Our results point to the existence of an unspanned volatility factor.

Keywords: Term structure models, Volatility, Forecasting, Kalman filtering, Yield curve

JEL Classification: C58, E43, G12

This version: June 28, 2021

1. Introduction

Affine term structure models have been the workhorse in the interest rate literature for decades and can successfully explain bond prices (see Duffee (2002), Christensen et al. (2011), and many others). However, it is questionable whether these models are able to explain yield volatilities, and hence reinvestment rate risk, which plays a crucial role in portfolio allocation, market timing, derivative pricing, and risk management. The literature has primarily focused on whether volatilities are spanned by interest rates, as implied by standard affine models. Collin-Dufresne et al. (2009) and Andersen and Benzoni (2010) conclude that they are not, while Jacobs and Karoui (2009) and Bikbov and Chernov (2009) reach mixed conclusions.

In this paper, we consider out-of-sample forecasting of yield volatility. Since future reinvestment rate risk matters for pricing and trading decisions, the yield curve should be sensitive to, and hence informative about, future yield volatility. Nevertheless, the forecasting perspective has so far received little attention in the yield volatility literature. The only exceptions are Collin-Dufresne et al. (2009) and Joslin and Konchitchki (2018), who consider fixed-window estimation, and do not compare the model based forecasts against time series benchmarks beyond a random walk (RW). In contrast, we update the parameters recursively according to investor's information set, and compare the model based volatility forecasts to leading time series benchmarks, in particular, the heterogeneous autoregressive (HAR) model of Corsi (2009) and the realized GARCH model of Hansen et al. (2012). Our focus is on whether the yield curve contains incremental information about future yield volatility, beyond that available from the time series of historical volatilities. In addition, we examine whether the time series based forecasters contain incremental information about future volatility, beyond the information that can be read off the yield curve. Further, we supplement the existing static analysis of the spanning issue with a dynamic out-of-sample assessment.

As a separate contribution, we use high-frequency intraday yield curves in the analysis, thus adding precision to volatility measurements. The past two decades have seen

important advances in high-frequency data analysis, primarily focusing on equities, e.g., Barndorff-Nielsen and Shephard (2002), Andersen et al. (2003), and Zhang et al. (2005). However, intraday data have seen little application in the interest rate literature. Notable exceptions are Faust et al. (2007), who use high-frequency data to study macroeconomic announcement effects, not volatility, and Andersen and Benzoni (2010) and Cieslak and Povala (2016), who use realized volatilities of interest rates based on a 10-minute sampling frequency, chosen to mitigate the effects of market microstructure noise. We overcome the latter issue in a more direct manner by relying on the pre-averaged realized variance estimator of Jacod et al. (2009), a noise-robust estimator of volatility with excellent empirical properties for various asset classes, as documented by Christensen et al. (2014). This choice allows us to use data at higher frequency. From US Treasury and Eurodollar futures prices, we construct yields with maturities 6 months, and 1, 5, and 7 years, at the 1-minute frequency. From these, we calculate a daily time series of realized volatility measures at each of the four maturities. We consider daily yield curve based forecasters of volatility over the subsequent month, constructed from either (i) affine term structure models, (ii) principal component analysis (PCA), or (iii) interest rate (yield and forward) spreads, estimated recursively in a standard daily yield panel. The alternative forecasters based on the time series of historical volatilities (RW, HAR, realized GARCH) serve as benchmarks.

We find that the yield curve contains superior information about future yield volatility relative to the time series of historical volatilities in terms of forecasting accuracy at longer maturities. The most accurate of the forecasts considered are generated by a common factor or interest rate spread approach at the 1, 5, and 7 year maturities, and by the HAR model at 6 months. This indicates that the yield curve contains some important information about future reinvestment rate risk that neither the term structure models nor the volatility history capture. Next, we carry out a specification test as in Cieslak and Povala (2016) and find that none of the affine or time series models considered subsumes all the relevant information about future volatility contained in the yield curve across all maturities. Extending this idea, we develop a test of whether past volatility forecast

errors are informative about future forecast errors, as would be the case under unspanned stochastic volatility (USV) in the sense of Collin-Dufresne and Goldstein (2002). We perform the test by extracting a common factor from the volatility forecast errors across maturities and find that it provides significant information about future forecast errors, thus indicating the presence of USV features in the data. Overall, we find that the yield curve contains incremental information about future volatility relative to the time series of historical volatilities at the long end of the curve, and that some of the information in the curve is not captured by standard affine models. At the short end, the time series models outperform yield curve based volatility forecasts.

We investigate the economic value of the forecasters to an investor in a portfolio allocation exercise, using a utility-based framework as in Bollerslev et al. (2018). From the results, the information about future interest rate risk contained in either the yield curve or historical volatility provides economic value to a risk averse investor in longer maturity bonds, whereas the RW forecast suffices at the short end of the curve, on utility grounds. A robustness analysis shows that higher-order PCA factors beyond the first three usually explaining yields in fact carry information about volatility. The portion of future volatility information in the yield curve that is not captured by standard affine term structure models is contained in these higher-order factors rather than in nonlinearities. Finally, a comparison with forecasts of future interest rate risk extracted from a wide cross section of coupon bonds confirms that the yield curve based forecasts are well represented by the daily panel estimates.

Our work on out-of-sample interest rate risk complements several strands of literature. Within the class of affine term structure models, conditional variances are affine in state variables. In standard versions, volatilities can be extracted from the cross section of interest rates. Under USV, some factors affect volatility, but not interest rates directly. Jacobs and Karoui (2009) examine the ability of three-factor affine models to fit conditional volatilities in-sample, using estimates from an EGARCH model to proxy for true volatility. They reach mixed conclusions, depending on the sample considered. Correlations between model-implied volatility and the proxy range between 60% and 75%

in US Treasury data, but are much lower in swap data, even negative for some maturities. Collin-Dufresne et al. (2009) consider affine models with and without a USV factor and conclude that volatility cannot be extracted from the cross section of interest rates based on swap data. Andersen and Benzoni (2010) find low R^2s for regressing future variances on PCA factors from the yield curve, and argue that they conduct a test of whether variances are spanned by bonds. The validity of this regression test of USV is questioned by Bikbov and Chernov (2009), who show that low R^2 is expected if yields are observed with error, a standard assumption when estimating term structure models. Bikbov and Chernov (2009), Joslin (2017), and Joslin and Konchitchki (2018) consider both swaps and swaptions and reject the restrictions needed for affine models to generate USV. We contribute to this literature by considering the spanning question in an out-of-sample framework, and providing evidence pointing to the existence of a USV factor.

The USV puzzle has recently led to the construction of more complex models, with a focus on the ability to fit volatilities and price swaptions. Feldhütter et al. (2016) propose a nonlinear model that is able to generate features consistent with USV. Cieslak and Povala (2016) construct a model with volatility factors driven by a Wishart process. Filipović et al. (2017) present a nonlinear model featuring USV that can price both bonds and swaptions successfully. None of these papers considers out-of-sample forecasting. It is beyond the scope of the present paper to include these more complex models, as they are very time-consuming to implement and reestimate recursively.

The paper is organized as follows. Section 2 describes the forecasting methods, including the term structure models, PCA and interest rate spread based forecasts, and time series models. Section 3 describes the high-frequency futures data and the construction of realized yield volatility measures. Section 4 presents the estimation method for the term structure models based on the Kalman filter. Section 5 contains the empirical analysis, and Section 6 the robustness analysis. Section 7 concludes. The Appendix contains additional material on data, models, estimation, and results.

2. Interest rate risk forecasters

Throughout, y_t^τ denotes the maturity τ yield at time t. In our empirical out-of-sample analysis, we consider daily forecasting of month-ahead interest rate risk by maturity. To assess the forecasts, we use the high-frequency data to calculate daily realized yield volatility, labeled V_t^τ (see Section 3.3). In line with the volatility literature, the target for the forecast built at t, based on investor's information set, is not the realized volatility at $t+h$, but instead the aggregated realized measure $V_{t+1|t+h}^\tau = \sum_{i=1}^h V_{t+i}^\tau$, a proxy for integrated volatility from t through $t+h$, or $IV_{t|t+h}^\tau$, say.[1] All forecasts considered are recursive, i.e., only data through t are used to construct the forecast of $V_{t+1|t+h}^\tau$, for each maturity τ and method considered, with h set to 22 trading days for month-ahead forecasting.[2] Methods utilizing the cross-sectional information in the yield curve are presented first, followed by methods relying exclusively on the information in the time series of historical volatilities.

2.1. Yield curve based volatility forecasters

We consider an $N \times T$ fixed-maturity panel, i.e., the yields at t are $y_t = (y_t^{\tau_1}, \ldots, y_t^{\tau_N})'$, $t = 1, \ldots, T$, and yields of all maturities can be used to generate volatility forecasts at a given maturity τ. For each candidate forecasting method, recursive volatility forecasts are constructed using the regression

$$V_{t+1|t+h}^\tau = \alpha^{\tau,h} + \beta^{\tau,h\prime} Z_t + u_{t+h}^{\tau,h}\,, \tag{1}$$

for fixed τ, h, with $\beta^{\tau,h}$ a $q \times 1$ vector of predictive coefficients, and Z_t the relevant information variable extracted from the yield panel and conditioning the forecast as of t for the given method, e.g., a q-vector of fitted PCA factors, or a conditional volatility forecast from an affine model. When constructing the forecast at t' of $V_{t'+1|t'+h}^\tau$, only data through t' are used to extract $Z_1, \ldots, Z_{t'}$ from the yield panel, then estimate Eq. (1) over

[1] If the yield follows the Itô process $dy_t^\tau = \mu_t^\tau dt + \sigma_t^\tau dW_t$, then $IV_{t|t+h}^\tau = \int_t^{t+h} (\sigma_s^\tau)^2\, ds$.

[2] Results for $h = 44$, forecasting two months ahead, are in the Appendix.

$t = 1, \ldots, t' - h$, and the forecast is $\hat{\alpha}^{\tau,h} + \hat{\beta}^{\tau,h\prime} Z_{t'}$. We consider methods for extracting Z_t based on affine term structure models, PCA, common factors, and risk premiums.

The recursive regressions in Eq. (1) are used in the construction of volatility forecasts, for two reasons. First, investor is not basing the forecast of $V^{\tau}_{t'+1|t'+h}$ exclusively on $Z_{t'}$ from the daily yields, having observed the history of realized volatilities through t'. Second, in the implementation, volatility forecasts based on Z_t include estimation error, whereas realized yield variance used as forecasting target in the assessment includes measurement error, thus leading to potential biases. Therefore, the Mincer and Zarnowitz (1969) type predictive regressions in Eq. (1) are used to leverage any indication in investor's information set that the centering or scale of Z_t from the given method do not line up well with subsequent realized volatilities. The main question is whether Z_t serves as a useful information variable, i.e., whether variation in this provides incremental predictive power. If, for example, a combination of the most recent realized volatility observations $V^{\tau}_{t'}, V^{\tau}_{t'-1}, \ldots$ provides the best forecast at t', as in the HAR model, then this feature will not be fully captured by the intercept and slope in Eq. (1), estimated over the full window $t = 1, 2, \ldots, t' - h$. It should be possible to improve on the resulting forecast using pure time series methods and historical volatilities, hence revealing that Z_t from the panel does not carry incremental information. Thus, the regression approach secures a level playing field for comparison of yield curve and time series based yield volatility forecasts.

In our empirical work, we use $N = 8$ maturities to extract Z_t from daily data, and volatility forecasts based on investor's information set are constructed using the recursive regressions in Eq. (1) for each of the four maturities τ for which we have high-frequency data.

2.1.a. Affine term structure models

In the affine class, the short rate $r_t = y_t^0$ is driven by some d-dimensional vector of state variables, X_t, such that $r_t = \delta_0 + \delta_1' X_t$, and the dynamics of X_t are

$$dX_t = \kappa(\theta - X_t)dt + \Sigma\sqrt{S(X_t)}dW_t\,, \tag{2}$$

where W_t is a d-dimensional Brownian motion under the physical measure $\mathbb{P}$, κ and Σ are $d \times d$ matrices, δ_1 and θ are $d \times 1$ vectors, and δ_0 is a scalar. In addition, $S(X_t)$ is a diagonal $d \times d$ matrix with each element affine in X_t,

$$[S(X_t)]_{ii} = \alpha_i + \beta_i' X_t\,, \tag{3}$$

with α_i a scalar and β_i a $d \times 1$ vector. Given a suitable market price of risk specification λ_t (see Appendix C), the dynamics of X_t are governed by an affine diffusion under the risk-neutral measure $\mathbb{Q}$, too,

$$dX_t = \tilde{\kappa}(\tilde{\theta} - X_t)dt + \Sigma\sqrt{S(X_t)}dW_t^{\mathbb{Q}}\,. \tag{4}$$

Following Duffie and Kan (1996), the yields are given by

$$y_t^\tau = \frac{A(\tau)}{\tau} + \frac{B(\tau)'}{\tau}X_t\,, \tag{5}$$

where A, B are solutions to the system of ordinary differential equations

$$\begin{aligned}
\frac{dA(\tau)}{d\tau} &= \tilde{\theta}'\tilde{\kappa}'B(\tau) - \frac{1}{2}\sum_{i=1}^{d}\left[\Sigma' B(\tau)\right]_i^2\alpha_i + \delta_0\,,\ A(0) = 0\,,\\
\frac{dB(\tau)}{d\tau} &= -\tilde{\kappa}'B(\tau) - \frac{1}{2}\sum_{i=1}^{d}\left[\Sigma' B(\tau)\right]_i^2\beta_i + \delta_1\,,\ B(0) = 0\,,
\end{aligned} \tag{6}$$

which can be solved either analytically or numerically.

By Eq. (5), the h periods ahead conditional variance of the maturity τ yield, given

information through t, is

$$V_{t,h}^{\tau} = Var_t(y_{t+h}^{\tau}) = \frac{1}{\tau^2} B(\tau)' Var_t(X_{t+h}) B(\tau) = b_0^{\tau,h} + b_1^{\tau,h\prime} X_t \,, \tag{7}$$

i.e., affine in the latent state variables. The precise forms of $b_0^{\tau,h}$ and $b_1^{\tau,h}$ are given in Appendix D.3. We consider affine models with d, the number of factors, from 1 to 3. Table 1 summarizes the models considered, including reference, $A_m(d)$ classification as in Dai and Singleton (2000), m being the number of factors conditioning variances, and market price of risk specification. In Appendix C, we provide a formal description of each model. We use Eq. (7) as the basis of our first yield curve based forecasters. Given the diversity of models considered, we are able to examine the impact of both the number of latent factors and stochastic (state dependent) volatility on the forecasting of future interest rate risk.

Insert Table 1 About Here

Implementation of Eq. (7) requires values for $b_0^{\tau,h}$, $b_1^{\tau,h}$, and X_t. We estimate the models recursively by quasi maximum likelihood (QML) using the Kalman filter and an expanding estimation window, allowing for additive measurement errors in yields (see Section 4 for details). To construct the forecast at t' of subsequent realized yield volatility $V_{t'+1|t'+h}^{\tau}$, only information through t' is used to generate $b_0^{\tau,h}$, $b_1^{\tau,h}$, and $X_1, \ldots, X_{t'}$, and hence the conditional volatility (or variance) estimates $\hat{V}_{t,h}^{\tau}$, $t = 1, \ldots, t'$, from Eq. (7), for each maturity τ, and fixed h. Next, Eq. (1) is estimated with $Z_t = \hat{V}_{t,h}^{\tau}$, $t = 1, \ldots, t' - h$, i.e., $q = 1$ in this case, and the final forecast of $V_{t'+1|t'+h}^{\tau}$ is $\hat{\alpha}^{\tau,h} + \hat{\beta}^{\tau,h\prime} \hat{V}_{t',h}^{\tau}$. Thus, even with correct model specification, the conditional variance itself is not an unbiased forecast of subsequent realized yield volatility. By the law of total variance, it includes the variance of the conditional mean yield, given both information through t' and integrated volatility through $t'+h$, thus inducing an upward bias.[3] The recursive regressions in Eq. (1) provide

[3] Write $\mathcal{F}_t$ for the information set. We have $Var(y_{t+h}^{\tau} \mid \mathcal{F}_t) = \mathbb{E}(Var(y_{t+h}^{\tau} \mid \mathcal{F}_t, IV_{t|t+h}^{\tau}) \mid \mathcal{F}_t) + Var(\mathbb{E}(y_{t+h}^{\tau} \mid \mathcal{F}_t, IV_{t|t+h}^{\tau}) \mid \mathcal{F}_t)$, where $Var(y_{t+h}^{\tau} \mid \mathcal{F}_t, IV_{t|t+h}^{\tau}) = IV_{t|t+h}^{\tau}$. Hence, $V_{t,h}^{\tau} = Var(y_{t+h}^{\tau} \mid \mathcal{F}_t)$ equals $\mathbb{E}_t(IV_{t|t+h}^{\tau}) \approx \mathbb{E}_t(V_{t+1|t+h}^{\tau})$ (as realized volatility converges rapidly to integrated volatility in high-frequency data) plus a positive bias stemming from conditional mean variation, $Var(\mathbb{E}(y_{t+h}^{\tau} \mid \mathcal{F}_t, IV_{t|t+h}^{\tau}) \mid \mathcal{F}_t)$. The latter is negligible for short horizons h (cf. Andersen et al. (2006)), but potentially important for our month-ahead forecasts. See also Appendix D on conditional mean variation in the Kalman filter.

a simple means of converting conditional variances into proper forecasts of future interest rate risk based on investor's information set.

2.1.b. PCA based forecasters

To examine whether or not the yield curve contains relevant information about future volatility and, if it does, whether the term structure models capture this information, we construct volatility forecasts using PCA factors based on the yield curve. We assume that the $N \times T$ yield panel is generated by a factor model,

$$y_t = \pi_0 + \pi_1' F_t + \varepsilon_t \,, \tag{8}$$

where F_t is a k-vector of common covariance-generating factors, $k < N$, π_1 is a $k \times N$ matrix of factor loadings, and π_0 represents de-meaning. We estimate F_t by PCA using the methodology of Bai and Ng (2002). There is some consensus in the literature that three factors can explain the variation in the yield curve (e.g., Litterman and Scheinkman (1991)). However, whether more factors can improve volatility forecasts, and hence whether different factors are important for explaining yields and for forecasting volatilities are open questions.

Recursive volatility forecasts are built using the fitted PCA factors $\hat{F}_t$ from Eq. (8) for Z_t in Eq. (1), for fixed τ, h, i.e., $q = k$ in this case. In the empirical analysis, we use $k = 3$ to 6 factors.

2.1.c. A common factor approach

For our third yield curve based forecaster, we consider a common factor approach, using volatility information across maturities to combine the maturity-specific forecasts from Section 2.1.b. The approach is similar to that used for forecasting bond risk premiums in Cochrane and Piazzesi (2005), based on forward rates, and in Ludvigson and Ng (2009), based on macro variables. Here, we construct a common factor for interest rate risk forecasting based on information from the yield curve.

Let $\tilde{F}_t$ denote some combination of fitted PCA factors from Eq. (8). We regress sub-

sequent variance averaged across maturities on $\tilde{F}_t$,

$$\frac{1}{4}\sum_{i=1}^{4} V_{t+1|t+h}^{\tilde{\tau}_i} = \gamma_0^h + \gamma_1^{h\prime}\tilde{F}_t + w_{t+h}^h\,, \tag{9}$$

for fixed h, where $\tilde{\tau}_1, \ldots, \tilde{\tau}_4$ are the four maturities for which we have high-frequency data. Recursive volatility forecasts are constructed using Eq. (1), for each maturity $\tau = \tilde{\tau}_i$, separately, with the fitted values $\tilde{V}_t$ from Eq. (9) for Z_t, i.e., $q = 1$ in this case. In $\tilde{F}_t$, we allow for selection from among the first six PCA factors from Eq. (8), and in robustness checks we allow squares, cubes, and interactions in the first three of these (see Section 6.1). Following Ludvigson and Ng (2009), the optimal combination of terms to include in $\tilde{F}_t$ is selected by minimizing the Bayesian information criterion (BIC) for Eq. (9). We consider both selection based on the initial estimation period, only, and recursive updating of the selection every period.

2.1.d. Risk premium based forecasters

Motivated by the literature on forecasting risk premiums in the bond market, we use interest rate spreads to form two additional yield curve based volatility forecasters. The first is the yield spread of Campbell and Shiller (1991). In this case, the explanatory variable for the maturity τ yield volatility is simply the difference between the maturity τ yield itself and the short rate, which we proxy by the CRSP 1-month T-bill rate, denoted $y_t^{1/12}$. Thus, volatility forecasts are constructed from Eq. (1), with $Z_t = y_t^\tau - y_t^{1/12}$, i.e., $q = 1$. The final risk premium based forecaster is based on the forward spread of Fama and Bliss (1987). We use the difference between the maturity τ forward rate, $f_t^\tau = y_t^\tau + (\tau - h)\left(y_t^\tau - y_t^{\tau-h}\right)/h$, and the short rate as explanatory variable. Thus, maturity τ yield volatility forecasts are constructed from Eq. (1), with $Z_t = f_t^\tau - y_t^{1/12}$.

2.2. Time series based yield volatility forecasters

As benchmarks, we consider three simple time series models, namely, the random walk (RW), the HAR model of Corsi (2009), and the realized GARCH model of Hansen et al. (2012). All three are widely used for volatility forecasting in equity markets. The RW

forecast of $V^\tau_{t+1|t+h}$ is simply $V^\tau_{t-h+1|t}$. The realized GARCH specification is chosen because it utilizes the information in the high-frequency data in forecasting, similarly to RW and HAR.

2.2.a. The HAR model

The HAR model is given by

$$V^\tau_{t+1|t+h} = \beta_0^{\tau,h} + \beta_D^{\tau,h} V_t^\tau + \beta_W^{\tau,h} V^\tau_{t-4|t} + \beta_M^{\tau,h} V^\tau_{t-21|t} + u^{\tau,h}_{t+h}\,. \tag{10}$$

The motivation for the cascade specification is the hypothesis on heterogeneous beliefs, according to which different investors react to information from the past day, week (5 trading days), and month (22 days). The model is estimated using nonoverlapping dependent variables, i.e., estimation given information through t' uses $t = t' - h, t' - 2h, \ldots$ in Eq. (10). Upon estimation, the forecast as of t' is given directly by Eq. (10), for $t = t'$.

2.2.b. Realized GARCH

Our final time series benchmark is the realized GARCH model. To accommodate mean reversion in yields, we consider a slightly extended model specification, given by

$$y_t^\tau - y^\tau_{t-h} = a^\tau + b^\tau y^\tau_{t-h} + \sqrt{h_t^\tau}\, z_t^\tau\,, \tag{11}$$

$$h_t^\tau = c^\tau + \sum_{i=1}^{p} d_i^\tau h^\tau_{t-ih} + \sum_{j=1}^{q} g_j^\tau V^\tau_{t-(j+1)h+1|t-jh}\,, \tag{12}$$

$$V^\tau_{t-h+1|t} = \xi^\tau + \phi^\tau h_t^\tau + \vartheta_1^\tau z_t^\tau + \vartheta_2^\tau((z_t^\tau)^2 - 1) + u_t^\tau\,, \tag{13}$$

where h_t^τ is the conditional variance of y_t^τ, given information through $t-h$, $z_t^\tau \sim \mathcal{N}(0,1)$, $u_t^\tau \sim \mathcal{N}(0, \sigma^2_{\tau,u})$, and suppressing the dependence of parameters on h. The standard model for equities, with returns (log-price changes) rather than yield changes on the left-hand side of Eq. (11), corresponds to the special case $b^\tau = 0$, and the extended model accommodates mean reversion in yields for $b^\tau \in (-1, 0)$ (at a rate corresponding to $\kappa = -\log(1+b^\tau)/h$ in continuous time, see Appendix D). For purposes of forecasting volatility

over the $h = 22$ day horizon, we consider a monthly specification, i.e., given information through t', the model is fit to yields and realized measures at $t = t', t' - h, t' - 2h, \ldots$, in analogy with the HAR model.

The model is estimated by maximum likelihood. We adopt the specification $p = 1$, $q = 2$ from Hansen et al. (2012). Upon estimation, the forecast as of t' is $\hat{\xi}^\tau + \hat{\phi}^\tau h^\tau_{t'+h}$. For comparison, we applied the standard model with either yields or yield changes on the left-hand side of Eq. (11), i.e., imposing $b^\tau = -1$ and 0, respectively. From the results, the extended mean-reverting realized GARCH model performs best for purposes of yield volatility forecasting, so we focus on this specification in our empirical work.[4]

3. Data description and construction of realized measure

3.1. Data description and cleaning

Our high-frequency data set is based on the three Treasury futures (5 years, 10 years, and long term) for the long end of the yield curve, and 3-month Eurodollar futures for the short end. The Treasury futures are traded on the Chicago Board of Trade (CBOT), and the Eurodollar futures on the Chicago Mercantile Exchange (CME). Data are obtained from CME Group. For liquidity reasons, we start our sample on January 2, 2000, and data run through October 31, 2016. Due to the opening hours in the early part of our sample, we consider a trading day from 7:20 am to 2:00 pm Eastern Time. All weekends and non-working days are excluded. This leaves a raw data set covering 4,207 days.

The Treasury futures have delivery dates in four different months during the year—March, June, September, and December. For each underlying bond maturity, we include the futures contract with highest liquidity, in most cases that with shortest term to delivery. For the Eurodollar futures, we include the two contracts with closest to 3 and 9 months to delivery, and convert the prices of these into yields of 6 months and 1 year to maturity. Section 3.2 describes the details of the procedure.

For each of the five contracts (three Treasury and two Eurodollar futures), we construct minute bars containing the first observation in each minute. If multiple observations on

[4]Results for the standard realized GARCH model are available from the authors on request.

the same contract have the same time stamp, we use the median price. We apply a sanity check, discarding observations at a distance more than ten times the absolute mean from the median of the previous ten observations. Finally, we exclude days without any observation for a whole hour for at least one of the five futures contracts. This eliminates 93 trading days, leaving 4,114 observation days in the final analysis data set. Appendix B describes the liquidity of the Treasury contracts.

3.2. Construction of intraday yield curves using futures data

We extract yield curves from futures prices in a two-step approach. In the first step, futures prices are converted into coupon bond prices. In the second, yield curves are extracted from the coupon bond prices. Our approach differs from that of Faust et al. (2007) mainly in two respects. First, we include more information on the long end of the curve, using all three Treasury futures, rather than only 10-year contracts. Second, following Dai et al. (2007), Andersen and Benzoni (2010), and others, we calibrate yield curves using the cubic spline method of Waggoner (1997), instead of the Nelson and Siegel (1987) method. In the following, we briefly outline the approach. More details on the first step are provided in Appendix A.1, and on the second in Appendix A.2.

Both Treasury and Eurodollar futures are considered in the first step. The underlying bond of a Treasury futures contract is hypothetical, and unknown until delivery. The seller chooses a bond from a delivery basket consisting of bonds meeting requirements from CME, implying that a Treasury futures involves two options: When to deliver, and which bond to deliver. We neglect the value of the first option by assuming that the delivery date is the first working day of the delivery month. Next, given the delivery date, we construct the delivery basket by combining the futures data with daily CRSP data on Treasuries, find the cheapest-to-deliver (CTD) bond in the basket thus constructed, and calculate the futures-implied coupon bond price.[5]

[5]Treasuries are non-convertible, and starting in 2000 avoids problems with the flower bonds issued until 1965, as these all matured by 1998. Callable bonds and notes were issued until 1985, but many of these subsequently repurchased by the Treasury and reissued as non-callable, although on a discretionary basis, without sinking fund provision. We drop the small number of remaining callable issues outstanding as of January 2, 2000.

The value of the Eurodollar futures at maturity is determined by the 3-month LIBOR rate. From the observed futures price, we derive a LIBOR rate with maturity date 3 months after the delivery date. LIBOR rates are afflicted with credit risk, and we deduce Eurodollar-implied government yields by assuming a constant credit spread between LIBOR and government yields. Daily LIBOR rates with maturities of 6 months and 1 year are obtained from the St. Louis Fed and compared with daily Gürkaynak et al. (2007) yields (see Section 3.4 for the construction) to estimate the spread. Finally, Eurodollar-implied government yields are converted into zero-coupon bond prices, so that units match with the coupon bonds.

In the second step, extracting yield curves from the futures-implied bond prices using cubic splines, an implication of our approach is that the maturities of bonds considered vary over the sample period, and therefore knot points vary, too. Due to the small number of cross-sectional observations (five futures contracts), we choose not to interpolate the calibrated curves between points of principal payments. The outcome of the procedure is a complete set of consecutive yield curves at the 1-minute frequency.

Yields are read off the calibrated curves at maturities τ reflecting the underlying maturities of the futures contracts, to ensure a sufficient amount of market based information around the relevant points along the curves. Initially, we consider $\tau = 0.5$, 1, 5, 7, and 15 years. The match to the maturities underlying the corresponding futures is by construction for the first two, and holds to a reasonable degree for the third, while $\tau = 7$ is included as the closest among standard maturities to the approximately 7.5 year average maturity of the Treasury note underlying the 10-year contract. The underlying of the long term Treasury futures is a hypothetical 6% bond. At expiration, a bond with maturity between 15 and 25 years must be delivered. When the interest rate is below 6%, conversion rates favor bonds with short maturities, i.e., the underlying bond is actually of maturity 15 years for our out-of-sample window, and hence $\tau = 15$ is considered as a candidate fifth maturity for our construction. To verify the quality of the resulting high-frequency yields at the five candidate τ-values, we match a daily frequency subsample of them with daily Gürkaynak et al. (2007) yields. For the latter, the number of bonds with maturity around

15 years outstanding at any given point in time in the CRSP data is rather limited, hence implying a low weight around this maturity in the calibration, and we find a relatively low correlation between the resulting yields and the daily subsample of ours at $\tau = 15$. Thus, as we cannot verify the quality of our fit at the fifth maturity, we henceforth restrict attention to the yields read off our high-frequency curves at $\tau = 0.5$, 1, 5, and 7 years, although we continue to use all five futures contracts in the construction of the curves.

At the four maturities retained, correlations between the end-of-day subset of futures-implied yields extracted from our 1-minute data and the daily calibrated yields range between 99.58% and 99.86% (see Table A.1 in the Appendix). In effect, we have created a high-frequency intraday version of the Gürkaynak et al. (2007) dataset up to maturity $\tau = 7$. Table 2 shows descriptive statistics on our yield data, by maturity. The term structure of interest rates is upward sloping on average over the sample period, and the term structure of volatilities downward sloping at longer maturities. Skewness and kurtosis are modest, especially at longer maturities. Figure 1 displays the evolution through time in yields, by maturity, along with a three-dimensional view of the evolution of yield curves through calendar time. The transition to the zero lower bound (ZLB) regime around 2008 has a strong impact, especially at the shorter maturities.

Insert Table 2 and Figure 1 About Here

3.3. Realized volatility measure

Access to high-frequency data on intraday yield curves enables us to estimate volatilies using realized measures. The market for the Treasury futures used in the construction of yield curves (Section 3.2) has undergone dramatic changes in liquidity during our sample period, in part due to the introduction of the electronic trading pit in 2004, and data are likely afflicted with market microstructure noise (see Appendix B). The high frequency econometrics literature proposes estimators of volatility that are robust to microstructure noise, e.g., the two time scales estimator of Zhang et al. (2005), the realized kernel approach of Barndorff-Nielsen et al. (2008), and the pre-averagered realized variance of Jacod et al. (2009). We deal with the issue by implementing the pre-averaging

estimator following Christensen et al. (2014), who show that this is successful in empirical analysis of volatility.

For each maturity τ, and on each day, write the intraday yield data as $y^{\tau}_{\frac{i}{n}}$, for $i = 0, 1, \ldots, n$. Pre-averaged yield changes are calculated by averaging yields in a local neighborhood consisting of K observations,

$$\bar{y}^{\tau}_{i} = \frac{1}{K}\left(\sum_{j=K/2}^{K-1} y^{\tau}_{\frac{i+j}{n}} - \sum_{j=0}^{K/2-1} y^{\tau}_{\frac{i+j}{n}}\right), \tag{14}$$

with K the nearest even integer to $\theta\sqrt{n}$, and θ a tuning parameter. The pre-averaged realized variance is then calculated as

$$V^{\tau}_{t} = \frac{n}{n-K+2}\frac{1}{K\psi_K}\sum_{i=0}^{n-K+1}(\bar{y}^{\tau}_{i})^2 - \frac{\hat{\omega}^2}{\theta\psi_K}, \tag{15}$$

with $\psi_K = (1 + 2K^{-2})/12$, and $\hat{\omega}^2$ an estimate of the noise variance,

$$\hat{\omega}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}\left(y^{\tau}_{\frac{i}{n}} - y^{\tau}_{\frac{i-1}{n}}\right)\left(y^{\tau}_{\frac{i-1}{n}} - y^{\tau}_{\frac{i-2}{n}}\right). \tag{16}$$

Given our one-minute sampling, the number of observations in a day, n, is 380, and we set $\theta = 1$. Since we are interested in volatility over the entire day, not only the hours for which we have high-frequency data, we add the squared overnight difference in yields, following Andersen and Benzoni (2010) and Bollerslev et al. (2018). The risk measure over the next month is simply $V^{\tau}_{t+1|t+h}$, $h = 22$, aggregating the pre-averaged realized variances over 22 trading days.

Table 3 shows descriptive statistics on the daily annualized pre-averaged realized yield volatilities over one day, $\sqrt{V^{\tau}_{t}}$, in Panel A, and one month, $\sqrt{V^{\tau}_{t+1|t+22}}$, in Panel B, in percent, by maturity. Although the volatility (square root) form of the measures facilitates interpretation, e.g., a unit mean corresponds to a one percent annual yield volatility, the raw (variance) form is used for forecasting, and descriptive statistics for this are reported in Panels C and D.

Insert Table 3 About Here

The term structure of volatilities (variances) exhibits a hump shape across the four maturities considered, with highest average at the $\tau = 1$ year maturity. The standard deviation (time series variation in volatility) shows a similar pattern. Skewness and excess kurtosis are highest at the medium maturities, $\tau = 1$ and 5 for the variance and volatility measures, respectively, and lowest at the long end of the curve, $\tau = 7$, where they essentially vanish for the one-month volatility measure.[6] On average, volatilities over one month are slightly larger than over one day, possibly reflecting that some conditional mean variation remains in the realized measures due to finite sampling frequency, cf. footnote 3, as well as Jensen's inequality for the square root measures (volatility spikes are included in 22 measures rather than one). As expected, other moments are smallest for the aggregated measures.

Figure 2 displays the evolution through time in daily annualized pre-averaged realized one-month yield volatility in percent, by maturity, corresponding to Table 3, Panel B, along with a three-dimensional view of the volatility surface.

Insert Figure 2 About Here

Volatilities rise quite dramatically for the shortest maturities during the financial crisis and the transition to the ZLB regime, and after the transition drop to a lower level than before.

3.4. Daily yield data

To enable estimation of the affine term structure models, as well as the PCA, common factor, and risk premium based forecasters, on a daily yield panel with more than four observations in the cross section,[7] we consider the estimated parameters provided by Gürkaynak et al. (2007). Their daily frequency dataset is extracted from a large set of coupon bonds, using the Svensson (1994) method. The continuously compounded yield

[6] Skewness and kurtosis are close to 0 and 3.

[7] Our high-frequency yield curves are constructed using five futures contacts, and four yields are retained in the resulting high-frequency panel, cf. Section 3.2. The time series models are estimated using realized measures (Section 3.3) based on the high-frequency panel.

at maturity τ is written as

$$y_t^\tau = \beta_0 + \beta_1 \frac{1 - e^{-\frac{\tau}{\theta_1}}}{\frac{\tau}{\theta_1}} + \beta_2 \left(\frac{1 - e^{-\frac{\tau}{\theta_1}}}{\frac{\tau}{\theta_1}} - e^{-\frac{\tau}{\theta_1}} \right) + \beta_3 \left(\frac{1 - e^{-\frac{\tau}{\theta_2}}}{\frac{\tau}{\theta_2}} - e^{-\frac{\tau}{\theta_2}} \right), \tag{17}$$

and estimates of $(\beta_0, \beta_1, \beta_2, \beta_3, \theta_1, \theta_2)$ are provided at daily frequency.[8] We follow Christensen et al. (2010) and consider $N = 8$ maturities in the daily cross sections, namely, 3 and 6 months, and 1, 2, 3, 5, 7, and 10 years.

4. Estimation

The affine term structure models from Section 2.1.a are estimated on daily data, cf. Section 3.4, using the Kalman filter. The basic measurement and transition equations are obtained by allowing for measurement error ε_{t+h}^τ in the yields in Eq. (5), and discretizing the state dynamics in Eq. (2), i.e., the state space model is given by

$$y_{t+h}^\tau = \frac{A(\tau)}{\tau} + \frac{B(\tau)'}{\tau} X_{t+h} + \varepsilon_{t+h}^\tau \,, \tag{18}$$

$$X_{t+h} = C_h + D_h' X_t + \eta_{t+h} \,, \tag{19}$$

where $\varepsilon_{t+h}^\tau \sim \mathcal{N}(0, H_{\tau,h})$, $\eta_{t+h} \sim \mathcal{N}(0, Q_{t,h})$, $h = 1$ for daily data and daily time index t, with expressions for A and B in Appendix C, and for C_h, D_h, and $Q_{t,h}$ in Appendix D. For the two Gaussian models, the standard linear filter applies. For the stochastic volatility models, Cox et al. (1985) (henceforth CIR) and AFNS_3, we apply the extended Kalman filter, approximating transitions by Gaussian distributions.

Upon estimation, conditional variance estimates are computed using Eq. (7), now with h indicating the forecasting horizon ($h = 22$ for month-ahead forecasting), and corrected for the bias stemming from measurement error in yields, ε_{t+h}^τ in Eq. (18), producing $\tilde{V}_{t,h}^\tau = b_0^{\tau,h} + b_1^{\tau,h\prime} X_{t+h} + H_{\tau,h}$. Recursive volatility forecasts are constructed using $\tilde{V}_{t,h}^\tau$ for Z_t in Eq. (1).

For the stochastic volatility models, we consider, in addition, a further extension to

[8] Available at http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html and updated daily.

the estimation method by including second moments in the filter, following Cieslak and Povala (2016) and Feldhütter et al. (2016). Based on Eq. (7), the additional measurement equation, beside Eq. (18), is

$$V_{t+h}^{\tau} = b_0^{\tau,h} + b_1^{\tau,h\prime} X_t + \tilde{u}_{t+h}^{\tau,h}\,, \tag{20}$$

with V_{t+h}^{τ} the pre-averaged realized yield variance from Eq. (15), $\tilde{u}_{t+h}^{\tau,h} \sim \mathcal{N}(0, \sigma_{\tau,h}^2)$, and $h = 1$ for the daily data estimation. The Gaussian term structure models imply that the variance of the yields is constant, and the filter extended with (20) is not applied to these models, although the predictive regression correction in Eq. (1) is. Indeed, the latter largely serves as a computationally simple alternative to the second moment extension, and we find in our empirical work that the basic recursive regressions in Eq. (1) based on investor's information set perform better for forecasting purposes than including Eq. (20) in the state space model.[9]

5. Empirical results

We consider the forecasting of yield volatility over the next month. The first set of estimates is based on the period January 2, 2000, through December 31, 2007, and the next 100 observations are used for recursive estimation of the predictive regressions in Eq. (1). This leaves 2,167 observations for the out-of-sample period, covering June 7, 2008, through October 31, 2016. This way, the out-of-sample window starts just before the transition to the ZLB regime. For PCA, risk premium based forecasters, and the HAR model, forecasts are not restricted to be positive. To ensure meaningful volatility forecasts, we apply a sanity filter, such that we do not forecast below the 2.5 percentile of the empirical distribution of observed realized variances.[10]

[9] Appendix D provides further details on estimation, including the Kalman filter and the second moment extension.

[10] The empirical distribution of observed variances is updated recursively, and forecasts below the 2.5 percentile replaced by this.

5.1. Statistical value of interest rate risk forecasts

To assess the yield volatility forecasts against a RW, we consider the R^2_{OoS} measure of Campbell and Thompson (2007),

$$R^2_{OoS} = 1 - \frac{\sum_{t=t_0+1}^{T-h}(V^{\tau}_{t+1|t+h} - \hat{V}^{\tau}_{t,h,\xi})^2}{\sum_{t=t_0+1}^{T-h}(V^{\tau}_{t+1|t+h} - \hat{V}^{\tau}_{t,h,RW})^2}\,, \tag{21}$$

where $\hat{V}^{\tau}_{t,h,\xi}$ is the forecast from model ξ, t_0 is the end of the initial estimation period, and T is the end of the sample. A positive R^2_{OoS} indicates more accurate volatility forecasts from model ξ than from the RW. The null hypothesis $R^2_{OoS} \leq 0$ is tested against the alternative $R^2_{OoS} > 0$ using a one-sided Diebold and Mariano (1995) test based on the Newey and West (1994) variance estimator with automatic lag selection of Andrews (1991).

Insert Table 4 About Here

Table 4 presents the resulting R^2_{OoS} statistics in percent and Diebold-Mariano p-values for all forecasters and maturities considered. From the first column of the table, none of the forecasters significantly improves over the RW at the shortest maturity, $\tau = 0.5$ (6 months). Only the HAR model generates positive R^2_{OoS}, and the p-value, at 0.39, is insignificant at conventional levels. This indicates that at the short end of the curve, neither cross-sectional yield curve information nor time series modelling adds predictive information about future volatility beyond the naive forecast. In contrast, at maturities $\tau = 1$ year and longer, R^2_{OoS} is generally positive, i.e., most methods produce more accurate forecasts than the RW, with p-values similar across methods, declining as maturity increases, and turning significant at the 10% and 1% levels at $\tau = 5$ and 7 years, respectively. Thus, information from either the yield curve or the time series of volatilities can be put together to improve volatility forecasts relative to the RW at longer maturities, but not at the short end of the curve.

A few comparisons within Table 4 are worth noting. The first panel shows results for the term structure models. Introducing stochastic volatility leads to improved forecasts for the three-factor model, but not for the one-factor model, i.e., R^2_{OoS} is higher for

$AFNS_3$ than for $AFNS_0$, but about as high or higher for the Vasicek model than for CIR. The $AFNS_3$ model performs about as well as Vasicek, but $AFNS_0$ does not. In general, parsimony can be rewarded in forecasting comparisons, as it reduces parameter uncertainty and the risk of in-sample overfitting, albeit at the expense of increased risk of model misspecification. The results suggest that the one-factor specifications strike a better balance between parsimony and misspecification than $AFNS_0$, and that allowance for stochastic volatility is required to justify the additional parameters in the three-factor specifications.

In general, the $AFNS_0$ implementation is quite unsuccessful. Therefore, the table also shows results for $AFNS_0$ without the predictive regression correction in Eq. (1), which for this model only turns out not to improve forecasting performance across all maturities, but R^2_{OoS} remains lower than for the one-factor Vasicek model.[11] Further, including second moments in the extended Kalman filter does not lead to an overall improvement in volatility forecasting for the $AFNS_3$ model, consistent with the notion that Eq. (1) provides an adequate account of the second moment history in investor's information set.[12]

The next panels in Table 4 show results for the PCA, risk premium, and common factor based forecasters. For the latter, the label "Initial" indicates that the selection of PCA factors is based on the initial estimation period, and "Recursive" that it is updated every period. All these methods generate R^2_{OoS} measures about as high or higher than the term structure models in the first panel, especially at the longest maturities, suggesting that the yield curve contains some information about future interest rate risk that the term structure models do not capture. For the PCA based forecasters, including more than three components actually makes forecasts deteriorate, presumably due to loss of parsimony. This is consistent with Litterman and Scheinkman (1991), who found that three factors provide a good description of the term structure.

The last panel in Table 4 shows results for the time series models. Although the

[11] Results for the other models without regression correction are available from the authors on request.

[12] Forecasting performance is dismal for the other models when including second moments. Results are available from the authors on request.

HAR model generates the highest R^2_{OoS} at $\tau = 0.5$ in the table, the cross section based forecasters (term structure models, PCA based, etc.) generate higher measures than HAR at $\tau = 1$ and 5, and the risk premium and common factor based forecasters at $\tau = 7$, too. The mean-reverting realized GARCH model similarly generates higher R^2_{OoS} than HAR at longer maturities, but not higher than the risk premium and common factor based forecasters.

For comparison, Table A.3 in the Appendix show results for volatility forecasts two months ahead ($h = 44$), laid out in the same manner as Table 4. Although p-values are higher, presumably due to higher noise-to-signal ratio at the longer forecasting horizon, the overall pattern is confirmed. At the $\tau = 5$ year maturity, the risk premium based forecasts (yield and forward spread) generate highest R^2_{OoS} across all methods, and are borderline significant at 10%, while AFNS_3, PCA, HAR, and realized GARCH all get p-values of 15% or lower. For month-ahead forecasting, Table 4, the risk premium based forecasts similarly generate highest R^2_{OoS} across methods at $\tau = 5$, and common factor based forecasts at $\tau = 1$ and 7, hence reinforcing the notion of information about future volatility in the yield curve. Finally, from Table A.3, Vasicek replaces HAR as the highest R^2_{OoS} forecast over two months at the shortest maturity, but does not improve significantly over RW ($p = 20\%$). Henceforth, we focus on the one-month forecasting horizon.

Overall, the results show that information from either the yield curve or the time series of volatilities can be used to improve month-ahead ($h = 22$) volatility forecasts over the RW at longer maturities, but not at the short end of the curve. Further, the yield curve contains information about future volatility at longer maturities that is not captured by the term structure models.

5.2. Extracting incremental information

Here, we address the question of whether the yield volatility forecasting equations based on the affine term structure models suffer from omitted variable bias, i.e., whether the

expression for future volatility in Eq. (7) should be expanded to

$$V_{t,h}^{\tau} = b_0^{\tau,h} + b_1^{\tau,h\prime} X_t + b_2^{\tau,h\prime} \tilde{X}_t \,, \tag{22}$$

for forecasting purposes. We consider three ways of extracting the incremental information variable $\tilde{X}_t$:

(i) By PCA factors from the yield curve.

(ii) Combining (i) with a factor extracted from the past volatility forecast errors from the analysis in (i).

(iii) By the realized volatility measure based on high-frequency data.

Since PCA factors capture the shape of the yield curve nonparametrically, improved forecasting performance (here, significance of $b_2^{\tau,h}$) in case (i) indicates that the yield curve contains incremental information relative to that in the affine term structure model considered. In (ii), significance indicates the existence of a factor explaining yield volatilities, but not yields, i.e., a USV case. In other words, significance of $b_2^{\tau,h}$ shows in case (i) that the term structure models do not capture all relevant information about future volatility available from the yield curve, and in case (ii) that the yield curve itself does not capture all relevant information. In (iii), significance indicates that historical volatility contains incremental information not captured by the term structure models.

In addition, in (i), we subject the time series models to the same test, i.e., we examine whether the same yield curve PCA information variable $\tilde{X}_t$ can be used to improve the volatility forecasts from the time series models. If so, this indicates that the yield curve contains incremental information about future volatility, relative to that contained in the volatility history itself. Conversely, we subject all other volatility forecasts to the test (iii), i.e., we investigate whether historical volatility contains incremental information about future volatility not captured by the term structure models, or by the yield curve itself, including the PCA and risk premium based methods.

5.2.a. The incremental information in the yield curve

For (i), we regress volatility forecast errors from model ξ on PCA factors from the yield curve,

$$V^{\tau}_{t+1|t+h} - \hat{V}^{\tau}_{t,h,\xi} = \phi_0^{\tau,h} + \phi_1^{\tau,h\prime}\hat{F}_t + u^{\tau,h}_{t+h}\,, \tag{23}$$

for fixed τ, h, with $\hat{F}_t$ the three leading fitted PCA factors at time t from Eq. (8), and test for joint significance of $\phi_1^{\tau,h}$ ($k = 3$ coefficients). Under the null, the model generating the forecast $\hat{V}^{\tau}_{t,h,\xi}$ subsumes the information content on future volatility available in the PCA yield curve factors.

Insert Table 5 About Here

Table 5 presents results from the specification test. At the 5% level, the Vasicek, CIR, and AFNS_3 models capture all relevant information in the PCA yield curve factors about future volatility at maturity $\tau = 5$. On the other hand, at shorter or longer maturities, the null is strongly rejected, i.e., these models do not capture all relevant information in the curve. Perhaps surprisingly, the results for AFNS_0 are opposite, as this model appears to capture information in the curve about future volatility at all maturities other than $\tau = 5$. Judging from the relatively poor performance of AFNS_0 in Table 4, the results for this model in Table 5 can reflect lack of power due to noisy estimation of Eq. (23). Thus, the indication is that the yield curve contains incremental information about future volatility beyond that captured by the affine models, which tend to be most informative at the intermediate maturity $\tau = 5$.

For the time series models, the evidence suggests that the yield curve contains incremental information about future volatility at the long end of the curve, relative to that contained in the historical volatility series. The null is rejected at $\tau = 7$ for both HAR and realized GARCH models.

Taken together, the results indicate that the time series models provide the best volatility forecasts at the short end of the curve, whereas the yield curve contains important incremental information about future volatility at the long end. Some of the information

in the yield curve is not captured by standard affine models, although they are relatively informative at the intermediate maturity $\tau = 5$.

5.2.b. Can past forecast errors predict future forecast errors?

In (ii), we examine whether a factor extracted from past yield volatility forecast errors can predict future forecast errors. An implication of USV is that it should be possible to extract at least one factor from volatility which is not related to the yields. We investigate this possibility by recursively extracting a factor from the fitted residuals from Eq. (23), say, $PC_{u,t}^{\tau,h}$, then testing for whether this contains significant information about future residuals in the regression

$$\hat{u}_{t+h}^{\tau,h} = \psi_0^{\tau,h} + \psi_1^{\tau,h} PC_{u,t}^{\tau,h} + \eta_{t+h}^{\tau}\,, \tag{24}$$

for fixed τ, h. Here, $\hat{u}_{t+h}^{\tau,h}$ is the residual from Eq. (23) at $t+h$, and $PC_{u,t}^{\tau,h}$ is extracted from the fitted residuals at t and earlier, across maturities τ, using PCA. Only one factor is included, due to the small number of volatilities in the cross section. The first 100 observations are used to initialize the factor. A significant coefficient $\psi_1^{\tau,h}$ indicates that a serially dependent USV factor is relevant for the forecast.

INSERT TABLE 6 ABOUT HERE

Table 6 shows results from estimation of Eq. (24). At 5%, evidence of an omitted factor arises at the longest maturity for the Vasicek and AFNS_0 models, and for all maturities except the intermediate $\tau = 5$ for the other models. Significant coefficients $\psi_1^{\tau,h}$ are negative in all cases, consistent with mean-reversion in the USV factor. As the test is conservative, in that the specified models are augmented with PCA factors fitted to past yield curves, cf. Eq. (23), the results are indicative of the existence of a latent volatility factor which is spanned neither by the parametric term structure models, nor nonparametrically by the yield curve.

5.2.c. The incremental information in historical volatility

In (iii), we investigate whether historical volatility contains information about future volatility not captured by the models considered, using a similar regression as in Section 5.2.a, but with the lagged realized measure based on high-frequency data as predictor (see Andersen and Benzoni (2010) for a related analysis). The specification is Eq. (23), with V_t^τ replacing $\hat{F}_t$, and we test for $\phi_1^{\tau,h} = 0$.

INSERT TABLE 7 ABOUT HERE

Results appear in Table 7. According to this test, historical volatility contains no incremental information relative to the term structure models or common factor based forecasters, as statistics are generally insignificant across maturities.[13] While these forecasters do not improve over the RW at short maturities, only at long, from the results based on $R^2_{OoS} > 0$ in Table 4, they do subsume the information from the time series, based on Table 7. The model based forecasts may to some extent benefit from the predictive regression correction in Eq. (1), which incorporates volatility information through the regression coefficients, but from investor's perspective, this is available historical volatility, and should be allowed for in the assessment of the practical value of these models. Further, from Table 7, PCA based forecasters are informative at long maturities, and risk premium based at $\tau = 5$, whereas historical volatility contains incremental information relative to these at short maturities, except when using six PCA factors. Combining with Table 5, the results show that at the short maturities, the PCA based forecasters provide incremental information relative to the term structure models, while historical volatility provides incremental information relative to PCA, but not relative to the term structure models, i.e., the relation is not transitive, and no dominant volatility forecast stands out at the short end. From Table 7, among the time series models, HAR subsumes the information in historical volatility across all maturities. Realized GARCH is relatively most informative at $\tau = 5$, but is missing some information contained in V_t^τ at other maturities

[13] Similar results are obtained for the term structure models if V_t^τ in the regression is replaced by the forecast from the HAR model, but the latter has incremental predictive power relative to the common factor forecasts at the two longest maturities.

τ, possibly due to recursive estimation and regime switching (to ZLB).

Overall, our specification tests reveal that the term structure of interest rates contains incremental cross-sectional information about future risk at the long end, relative to the time series of historical volatilities, although some of the information is not captured by the affine models. At the short end, the time series contains information about future volatility only captured in the model-free (PCA) approach to the yield curve if using relatively many (six) factors. The affine models and the common factor approach do subsume this information content in historical volatility, but do not further improve over the RW forecast at the short end. Finally, we find evidence pointing to the existence of a USV factor in the data.

5.3. Economic value of interest rate risk forecasts

So far, the analysis has focused on statistical measures of predictive ability. We next examine whether the volatility forecasting methods considered generate utility in a portfolio allocation framework, following Bollerslev et al. (2018). To this end, the analysis is translated from the level of yields to returns. We consider an investor purchasing a zero-coupon bond of maturity $\tau + h$ at time t and selling the bond at $t + h$. Let $\tilde{r}_{t+h}$ denote the log return from this trading strategy. The relation between returns and yields is

$$\tilde{r}_{t+h} = -\tau y^{\tau}_{t+h} + (\tau + h) y^{\tau+h}_{t} \,, \tag{25}$$

so the conditional variance of the return as of t is given by

$$Var_t(\tilde{r}_{t+h}) = \tau^2 Var_t(y^{\tau}_{t+h}) \,, \tag{26}$$

i.e., depending on maturity and the yield volatility forecast we consider. Investor is assumed to have mean-variance preferences and access to a risk-free as well as a risky asset, the latter being the zero-coupon bond with time-varying volatility. Assuming a constant Sharpe ratio, investor's utility depends only on the variance of the risky asset. We consider maturities 7, 13, 61, and 85 months, so that the analysis depends on the

forecasts from the previous sections.

Let w_t be the portfolio weight allocated to the risky bond and $1 - w_t$ the allocation to the risk-free asset with return r_t^f. The return to the portfolio at time $t + h$ is

$$r_{t+h} = r_t^f + w_t rx_{t+h}\,, \tag{27}$$

where $rx_{t+h} = \tilde{r}_{t+h} - r_t^f$ is the excess return to the risky asset. Since r_t^f is common across all forecasters, we only consider utility in terms of excess return. Expected utility per unit of wealth is given by

$$\mathcal{U}_{t+h} = w_t \mathbb{E}(rx_{t+h}) - \frac{1}{2}\gamma w_t^2 Var_t(rx_{t+h})\,, \tag{28}$$

where γ is relative risk aversion, and $Var_t(rx_{t+h}) = Var_t(\tilde{r}_{t+h})$ from Eq. (26). The optimal weight w_t^* is then

$$w_t^* = \frac{1}{\gamma}\frac{\mathbb{E}_t(rx_{t+h})}{Var_t(rx_{t+h})}\,, \tag{29}$$

which, given the constant Sharpe ratio, $SR = \mathbb{E}_t(rx_{t+h})/\sqrt{Var_t(rx_{t+h})}$, becomes

$$w_t^* = \frac{1}{\gamma}\frac{SR}{\sqrt{Var_t(rx_{t+h})}}\,. \tag{30}$$

By Eqs. (27) and (30), the volatility target sought by the investor is $w_t^*\sqrt{Var_t(rx_{t+h})} = SR/\gamma$. If the forecasted volatility $\tau\sqrt{Var_t(y_{t+h}^\tau)}$ exceeds this target, investor will place only a portion of wealth in the risky asset, $w_t < 1$, and save the remainder in the risk-free asset. Conversely, when the forecasted volatility falls short of the target, investor will take a geared position in the risky asset, $w_t > 1$, financed by borrowing at the risk-free rate. Thus, investor follows a volatility timing strategy.

To examine the utility gains from the various forecasters, let $\mathbb{E}_t(\cdot)$ denote the conditional expected value from the true model, $\mathbb{E}_t^\xi(\cdot)$ the conditional expected value from model ξ, and similarly for conditional variances. Assuming that investor uses model ξ for portfolio selection, the optimal weight becomes $w_{t,\xi}^* = SR/\left(\gamma\tau\sqrt{Var_t^\xi(y_{t+h}^\tau)}\right)$. The

expected utility per unit of wealth can then be expressed as

$$\mathcal{U}_{t+h} = \frac{SR^2}{\gamma}\left(\frac{\sqrt{Var_t(y^\tau_{t+h})}}{\sqrt{Var^\xi_t(y^\tau_{t+h})}} - \frac{1}{2}\frac{Var_t(y^\tau_{t+h})}{Var^\xi_t(y^\tau_{t+h})}\right). \tag{31}$$

If a model is able to perfectly predict conditional variance, then investor's utility is $SR^2/(2\gamma)$. Following Bollerslev et al. (2018), we set the annualized Sharpe ratio to 0.4. The risk aversion parameter, γ, is set to 5, following Sarno et al. (2016) and Gargano et al. (2017) from the bond return prediction literature. This corresponds to an annualized volatility target of 8%.

Insert Table 8 About Here

To evaluate predictors, we simply average realized utility. Table 8 reports the results. At the two shortest maturities, the RW generates highest realized utility across all volatility forecasting methods. At the 5 year maturity, investor is better off ignoring the RW forecast and investing in the risk-free asset, which generates utility zero. Even higher utility is achieved at the two longest maturities using either HAR, realized GARCH, or risk premium based forecasters, all of which improve significantly over RW, as does CIR including second moments at $\tau = 5$. The yield spread based forecast, HAR, and realized GARCH are essentially tied for highest utility at the two longest maturities, along with Vasicek at the longest, $\tau = 7$, where PCA provides value, too. Among the term structure models, CIR generates highest utility at maturities six months and one year, CIR including second moments at five years, and Vasicek at the longest maturity. The CIR, AFNS_0, and AFNS_3 models perform better than the PCA based forecasters across all but the longest maturity. Apparently, at the two longest maturities, the inclusion of second moments is better justified on utility based rather than statistical grounds, as it generates higher realized utility for CIR, although not for AFNS_3.

Overall, the analysis shows that the information about future reinvestment rate risk contained in both the yield curve (Vasicek, PCA, interest rate spreads) and historical volatility (HAR, mean-reverting realized GARCH) provides economic value to a risk averse investor in longer maturity instruments, not in short.

6. Robustness

This section addresses the robustness of the results in Section 5. First, we consider whether interest rate risk forecasts are improved by including nonlinear terms in PCA factors. Next, we examine whether there is a trade-off between explaining yields and forecasting volatility. Finally, we assess the information about future yield volatility in a wide cross section of coupon bond prices.

6.1. Nonlinearities and interest rate risk

Feldhütter et al. (2016) find that including nonlinearities via quadratic and cubic terms as well as interactions in factors increases in-sample explanatory power for realized variance. In our out-of-sample framework, we find in Table 4 that using the three leading factors $\hat{F}_t$ extracted from the yields at time t based on Eq. (8) suffices for the PCA based forecasts, i.e., adding factors beyond three does not increase accuracy. Here, we examine whether including nonlinearities in the first three PCA factors improves forecasts, both for the common factor approach and the maturity-specific[14] forecasts. As in the linear common factor approach, Eq. (9), the selection of terms to include is based on the BIC, either for the initial estimation period, or updated recursively.

Insert Table 9 About Here

Results appear in Table 9. For the common factor approach, selecting nonlinear terms based on the initial estimation period, only, generates more accurate forecasts than recursively updating the selection.[15] This finding points to a regime switch in which the ZLB regime is more similar to the initial estimation period than to the transition period, i.e., forecasts into the ZLB period are harmed by updating the selection during the transition. The same applies for the maturity-specific forecasts at the two longest maturities. Further, at these maturities, the common factor based forecasts are more precise than the maturity-specific. At the two shortest maturities, the maturity-specific

[14] In that no common factor based on Eq. (9) is used.

[15] A similar phenomenon is observed for the linear common factor based forecasts in Table 4, although only at the two shortest maturities.

forecasts with recursive selection of nonlinear terms are about as precise as the best common factor based forecasts. However, only common factor based forecasts improve significantly over the RW, and only at the long end of the curve, $\tau = 7$, although results are close to significant at $\tau = 5$, too (significant at 10% with initial selection). In the linear case, Table 4, similar (borderline) significant results at ($\tau = 5$ and) $\tau = 7$ were found not only for common factor based approaches, but for the maturity-specific (PCA based), too. Thus, including nonlinear terms makes maturity-specific forecasts lose significance at the long end, although the R^2_{OoS} measure itself is improved at the short end with recursive updating. Including nonlinear terms in the common factor approach makes the measure improve in the case with initial selection, but deteriorate at the long end in the recursive case, hence reinforcing the impression of a regime shift.

Insert Figure 3 About Here

Figure 3 shows yield volatility forecast errors over time, by maturity, for selected forecasting methods. Cumulative squared errors for a given forecasting method are subtracted from those for the RW, so that an increasing curve indicates better forecasting than by RW, and vice versa. In the common factor case, the improvement in the approach with initial selection of nonlinear terms relative to recursive updating occurs around 2009 for the two shortest maturities, consistent with the ZLB transition driving the phenomenon. For comparison, the figure shows results for Vasicek, CIR, and HAR, too. Most methods exhibit jumps in forecasting accuracy, moving into the ZLB regime, although the effect materializes later at longer maturities. It could be expected that CIR (with state-dependent volatility) would handle the entire ZLB period better than Vasicek, but this is only confirmed at the long end of the curve.[16] Historical volatility (HAR) forecasts better than the curve-based methods at the shortest maturity, but not at longer, where the common factor based forecasts dominate. Improvements relative to the RW continue steadily over calendar time at $\tau = 7$, especially for the yield curve based methods, consistent with incremental volatility information at the long end of the curve, whereas results stabilize through the ZLB period at shorter maturities.

[16] During the transition, around 2009, it holds at the short end, too.

Summing up, the results show that, first, allowing for nonlinearities does not generate an overall improvement in yield volatility forecasting. Second, the regime switch around the transition to the ZLB is confirmed. Following this, the curve-based methods improve over historical volatility at the long end.

6.2. Trade-off between explaining yields and forecasting volatility

The finding in Table 4 that using the first three PCA factors $\hat{F}_t$ from Eq. (8) generates more accurate yield volatility forecasts than using the first four, five, or six factors raises the additional question of whether using exactly these three is optimal, or whether some of the remaining factors beyond the first three are more informative about future yield volatility. Further, if volatility forecasts are indeed improved by using some other combination of factors, the question arises whether the optimal combination is common across maturities.

To address these issues, we consider a regression of the type in Eq. (1), with Z_t representing a combination of three of the first six PCA factors, i.e., not necessarily the first three. We consider all 20 possible combinations.

INSERT TABLE 10 ABOUT HERE

Table 10 presents the R^2_{OoS} results. As for the PCA based forecasts in Table 4, all combinations of factors significantly improve forecasting performance relative to the RW at level 1% at the longest maturity, $\tau = 7$, and at 10% at $\tau = 5$, whereas no combination outperforms the RW at the short end of the curve. For none of the maturities is volatility forecasting based on the three leading PCA factors (last line in the table) among the best choices. It is not among the worst specifications, with four or more combinations at each maturity performing worse (except only two at $\tau = 5$). Still, the evidence is clearly that using the first three factors is not optimal for purposes of forecasting yield volatilities. Further, the optimal combination of factors for volatility forecasting is not common across maturities, but it is close. At the long end of the curve, the best forecasts are obtained by using factors 3, 5, and 6, and this combination is among the best across all maturities. At $\tau = 5$, only two very similar combinations, $(3, 4, 6)$ and $(3, 4, 5)$, generate slightly higher R^2_{OoS} than $(3, 5, 6)$. At $\tau = 2$, only three of 20 combinations are better, and at $\tau = 1$ only

a single, similar one, $(2, 3, 6)$. The importance of including the sixth factor for purposes of yield volatility forecasting is consistent with the finding from Table 7 that only using six PCA factors subsumes the information about future yield volatility contained in historical volatility. The deterioration in performance by adding factors beyond the leading three in Table 4 indicates the importance of parsimony, i.e., although higher-order factors are informative about volatility, this is only revealed by forecasting using these factors alone.

The results indicate a trade-off between explaining yield levels and forecasting interest rate risk. The fourth and higher PCA factors explain future volatilities more than current yield curves, which to a large extent are explained by the first two factors, level and slope. The third factor, curvature, plays a role for both curves and volatilities.

6.3. Interest rate risk forecasts based on a cross section of coupon bond prices

Our analysis relies on panel data on yields constructed from high-frequency futures prices, with five price observations in the cross section dimension, for the time series methods, and on the daily panel with $N = 8$ yields in the cross section for the curve-based methods. Here, we investigate the possibility that the full cross section of observed coupon bond prices at any point in time carries more information about future interest rate risk. We consider daily CRSP data on quoted prices of bonds with maturity between 3 months and 10 years, adjusted for accrued interest, and excluding callable issues. The price of a bond promising M semi-annual payments of $C/2$ at $\tau = (\tau_1, \ldots, \tau_M)$ is represented as

$$P_t(\tau, C) = \frac{C}{2}\frac{\tau_1 - t}{1/2}B_t^{\tau_1} + \sum_{j=2}^{M}\frac{C}{2}B_t^{\tau_j} + 100B_t^{\tau_M}\,, \tag{32}$$

with $B_t^{\tau_j} = \exp(-\tau_j y_t^{\tau_j})$ the price of a zero-coupon bond of maturity τ_j, and $(\tau_1 - t)/(1/2)$ the fraction of the first coupon receivable, reflecting the share of the half-year interval between coupons over which the coupon bond is held. Bond prices $P_t^{i,obs}$ are assumed to be observed with measurement error,

$$P_t^{i,obs} = P_t(\tau^i, C^i) + \varepsilon_t^i\,, \tag{33}$$

with $P_t(\cdot)$ the model-implied price from Eq. (32), (τ^i, C^i) contractual terms for bond i, and the measurement error ε_t^i Gaussian with variance σ_t^2, independently across observations $i = 1, \ldots, N_t$ available at t. Write P_t^{obs} for the N_t-vector of observed coupon bond prices, Θ_t for the $\mathbb{Q}$-parameters of the pricing model, augmented with the short rate r_t, considered latent and backed out from the cross section, and $P(\Theta_t)$ for the N_t-vector of model-implied prices. Estimation is by cross-sectional nonlinear regression or QML,

$$\hat{\Theta}_t = \arg\max_{\Theta_t} \left(-\frac{N_t}{2} \log \sigma_t^2 - \frac{1}{2\sigma_t^2} (P_t^{obs} - P(\Theta_t))'(P_t^{obs} - P(\Theta_t)) \right) . \tag{34}$$

The cross-sectional estimate $\hat{\Theta}_t$ does not depend on σ_t^2. Figure A.2 in the Appendix shows the evolution over time in N_t, ranging from a low of 85 in the early part of the sample to a high of 272 towards the end.

For purposes of yield volatility forecasting, $\mathbb{Q}$-parameters do not suffice, as $V_{t,h}^{\tau}$ in Eq. (7) depends on $\mathbb{P}$-parameters through the coefficients $b_0^{\tau,h}$ and $b_1^{\tau,h}$. For concreteness, we focus on the CIR model, i.e., Eq. (34) is the estimator considered by Brown and Dybvig (1986), who examined the consistency between the cross-sectional estimate of the short rate volatility parameter σ in the square root process and a time series estimate of this, based on the sample variance of estimated short rates across calendar time. This comparison was relevant because σ is common across $\mathbb{Q}$ and $\mathbb{P}$ in the continuous-time model (see Appendix C.2). Instead, since we consider volatility forecasting at longer maturities, we estimate the market price of risk λ (cf. Table 1) from a time series of short rates calibrated period by period in Eq. (34). Specifically, we consider an Euler discretization of the CIR process,

$$r_{t+1} - r_t = \kappa(\theta - r_t) + \sigma\sqrt{r_t}\varepsilon_{t+1} , \tag{35}$$

following Chan et al. (1992), who focused on the $\mathbb{P}$-parameter estimates resulting from time series analysis of Eq. (35). Given the estimated $\mathbb{Q}$-parameters from Eq. (34), we instead express Eq. (35) as a function of λ. Since r_{t+1} conditional on r_t is Gaussian under the discretization, we estimate λ by QML. For forecasts at t, we use $\mathbb{Q}$-parameters from

Eq. (34) at t, along with λ estimated from a short time series $t-10,\ldots,t$ in Eq. (35). An alternative would be to estimate all three $\mathbb{P}$-parameters in a longer time series, and construct final estimates of all four parameters (including λ) by optimal weighting of estimates from the sequential cross-sectional and time series regressions, following Andreasen and Christensen (2015). Rather than pursuing this approach, we focus on the short time series of fitted r_t for setting λ, to retain the cross-sectional nature of the approach.

Insert Table 11 About Here

Table 11 shows the resulting R^2_{OoS} statistics. For comparison, results are shown both for forecasts based purely on cross sectionally estimated $\mathbb{Q}$-parameters and short rates, and the combined approach described. The latter generates the most accurate forecasts of the two at the long end, $\tau = 7$, confirming that the current risk premium carries future volatility information, and consistent with the relatively strong performance of the risk premium based forecasters in Table 4. At shorter maturities, the $\mathbb{Q}$-based forecasts generate slightly higher R^2_{OoS}, presumably due to parsimony. The improvement over the RW turns significant at 10% and 5% for $\tau = 5$ and 7, respectively, broadly in line with the pattern for curve based and time series models in Table 4, although significance at 1% is generally achieved at the long end in the latter case. Further, volatility forecasts based on the cross section of coupon bonds do not generate as high R^2_{OoS} as the risk premium based forecasts in Table 4, and not higher than PCA, either.

Overall, the robustness analysis confirms that the daily yield panel captures the cross-sectional information adequately. Further, more of the future volatility information in the curve not captured by standard term structure models appears to be contained in higher-order factors than in nonlinearities.

7. Concluding remarks

Our results show that the assessment of future interest rate risk is a complicated affair. Agents with a primary interest in the short end of the curve, such as market timers, should base their forecasts on historical volatility, with a focus on recent periods, as in the

formal time-varying volatility models. If the long end is important, e.g., for immunization of assets and liabilities, or valuation of capital assets, then incremental information is available by looking across maturities along the curve, and more so than implied by formal term structure models, which fare best at intermediate maturities, around five years. Some of the curve-based information about future volatility is distinct from that explaining the level of yields, and is contained in higher-order factors rather than nonlinearities. Based on a simple portfolio exercise, the information about future reinvestment rate risk in either the yield curve or the volatility history is of economic value to a risk averse investor in longer-term instruments, whereas reliance on the naive forecast (current volatility) suffices at the short end, on utility grounds. Finally, our results point to the existence of a latent volatility factor, unspanned by yields.

We have emphasized that when leveraging volatility information drawn from the yield curve, investor should not ignore historical volatility in the information set. That the resulting picture is diverse is perhaps not surprising. Investor is, after all, gazing into the crystal ball. Clearly, mimicking this situation requires a recursive, out-of-sample approach, to avoid admitting the artificial investor the benefit of hindsight.

References

Andersen, T. G. and L. Benzoni (2010). Do bonds span volatility risk in the US Treasury market? A specification test for affine term structure models. *Journal of Finance 65*(2), 603–653.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. In G. Elliot, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 778–878. Amsterdam: North-Holland.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andreasen, M. M. and B. J. Christensen (2015). The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models. *Journal of Econometrics 184*(2), 420–451.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica 59*(3), 817–858.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B 64*(2), 253–280.

Bikbov, R. and M. Chernov (2009). Unspanned stochastic volatility in affine models: Evidence from Eurodollar futures and options. *Management Science 55*(8), 1292–1305.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies 31*(7), 2729–2773.

Brown, S. J. and P. H. Dybvig (1986). The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates. *Journal of Finance 41*(3), 617–630.

Campbell, J. Y. and R. J. Shiller (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies 58*(3), 495–514.

Campbell, J. Y. and S. B. Thompson (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

Chapman, D. A., J. B. Long Jr, and N. D. Pearson (1999). Using proxies for the short rate: When are three months like an instant? *Review of Financial Studies 12*(4), 763–806.

Chen, R.-R. and L. Scott (2003). Multi-factor Cox-Ingersoll-Ross models of the term structure: Estimates and tests from a Kalman filter model. *Journal of Real Estate Finance and Economics 27*(2), 143–172.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics 164*(1), 4–20.

Christensen, J. H., J. A. Lopez, and G. D. Rudebusch (2010). Can spanned term structure factors drive stochastic yield volatility. Working paper, Federal Reserve Bank of San Francisco.

Christensen, K., R. C. Oomen, and M. Podolskij (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics 114*(3), 576–599.

Cieslak, A. and P. Povala (2016). Information in the term structure of yield curve volatility. *Journal of Finance 71*(3), 1393–1436.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Re-*

view 95(1), 138–160.

Collin-Dufresne, P. and R. S. Goldstein (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *Journal of Finance 57*(4), 1685–1730.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2009). Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics 94*(1), 47–66.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–408.

Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance 55*(5), 1943–1978.

Dai, Q., K. J. Singleton, and W. Yang (2007). Regime shifts in a dynamic term structure model of US Treasury bond yields. *Review of Financial Studies 20*(5), 1669–1706.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance 57*(1), 405–443.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*(4), 379–406.

Fackler, P. (2000). Moments of affine diffusions. Working paper, North Carolina State University.

Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review 77*(4), 680–692.

Faust, J., J. H. Rogers, S.-Y. B. Wang, and J. H. Wright (2007). The high-frequency

response of exchange rates and interest rates to macroeconomic announcements. *Journal of Monetary Economics 54*(4), 1051–1068.

Feldhütter, P., C. Heyerdahl-Larsen, and P. Illeditsch (2016). Risk premia and volatilities in a nonlinear term structure model. *Review of Finance 22*(1), 337–380.

Filipović, D., M. Larsson, and A. B. Trolle (2017). Linear-rational term structure models. *Journal of Finance 72*(2), 655–704.

Fisher, M., D. Nychka, and D. Zervos (1995). Fitting the term structure of interest rates with smoothing splines. Working paper, Federal Reserve Board.

Gargano, A., D. Pettenuzzo, and A. Timmermann (2017). Bond return predictability: Economic value and links to the macroeconomy. *Management Science 65*(2), 508–540.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007). The US Treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*(8), 2291–2304.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

Jacobs, K. and L. Karoui (2009). Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets. *Journal of Financial Economics 91*(3), 288–318.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications 119*(7), 2249–2276.

Joslin, S. (2017). Can unspanned stochastic volatility models explain the cross section of bond volatilities? *Management Science 64*(4), 1707–1726.

Joslin, S. and Y. Konchitchki (2018). Interest rate volatility, the yield curve, and the macroeconomy. *Journal of Financial Economics 128*(2), 344–362.

Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income 1*(1), 54–61.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*(12), 5027–5067.

Mincer, J. A. and V. Zarnowitz (1969). The evaluation of economic forecasts. In J. A. Mincer (Ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pp. 3–46. Cambridge: NBER.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*(4), 631–653.

Sarno, L., P. Schneider, and C. Wagner (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance 37*, 247–267.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. NBER Working Paper w4871.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.

Waggoner, D. (1997). Spline methods for extracting interest rate curves from coupon bond prices. Working paper, Federal Reserve Bank Atlanta.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

Table 1: Affine models

This table presents the reference, notation of Dai and Singleton (2000), and market price of risk specification for the affine term structure models considered.

Model	Reference	$A_m(d)$-notation	Market price of risk
Vasicek	Vasicek (1977)	$A_0(1)$	λ
CIR	Cox et al. (1985)	$A_1(1)$	$\lambda/\sigma\sqrt{r_t}$
AFNS_0	Christensen et al. (2011)	$A_0(3)$	$(\Sigma\sqrt{S(X_t)})^{-1}(\kappa(\theta - X_t) - \tilde{\kappa}(\tilde{\theta} - X_t))$
AFNS_3	Christensen et al. (2010)	$A_3(3)$	$(\Sigma\sqrt{S(X_t)})^{-1}(\kappa(\theta - X_t) - \tilde{\kappa}(\tilde{\theta} - X_t))$

Table 2: Descriptive statistics on high-frequency yields

This table presents the mean, standard deviation, skewness, and kurtosis for the end-of-day subset of our high-frequency annualized yields in percent, by maturity. The sample spans the period from January 2, 2000, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Mean	1.76	1.80	2.77	3.44
Std.	1.98	1.96	1.55	1.28
Skewness	0.96	0.90	0.42	0.15
Kurtosis	2.52	2.48	2.13	2.07

Table 3: Descriptive statistics on pre-averaged realized yield volatility

This table presents the mean, standard deviation, skewness, and kurtosis of the daily annualized pre-averaged realized yield volatilities, by maturity. Statistics for the square root (volatility) form are shown for one-day measures in Panel A, and for one-month ($h = 22$) measures in Panel B, both in percent. Statistics for the raw (variance) form are shown for one-day measures in Panel C, and for one-month ($h = 22$) measures in Panel D. The sample spans the period from January 2, 2000, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
	Panel A: One-day volatility, $\sqrt{V_t^\tau}$			
Mean	0.68	1.01	0.37	0.30
Std.	0.80	1.07	1.00	0.57
Skewness	3.61	4.11	6.71	2.54
Kurtosis	24.69	47.36	77.69	9.95
	Panel B: Month-ahead volatility, $\sqrt{V_{t+1\|t+22}^\tau}$			
Mean	0.90	1.28	0.84	0.77
Std.	0.60	0.82	0.70	0.27
Skewness	2.16	2.06	2.86	0.17
Kurtosis	9.56	11.96	16.06	2.95
	Panel C: One-day variance, V_t^τ			
Mean	1.09	2.19	1.13	0.41
Std.	3.95	9.34	9.31	1.27
Skewness	11.65	31.61	22.96	6.93
Kurtosis	187.4	1393.0	651.7	87.8
	Panel D: Month-ahead variance, $V_{t+1\|t+22}^\tau$			
Mean	1.16	2.32	1.20	0.44
Std.	1.94	3.93	2.89	0.35
Skewness	4.85	7.11	6.93	1.33
Kurtosis	34.19	72.69	61.07	4.93

Table 4: Out-of-sample R^2 for month-ahead yield volatility forecasting
This table displays R^2_{OoS} measures in percent relative to a RW for all forecasting methods and maturities. In parentheses asymptotic p-values for a one-sided Diebold-Mariano test using the Newey-West variance estimator with automatic lag selection of Andrews (1991). For common factor based forecasters, the label "Initial" indicates that the selection of PCA factors is based on the initial estimation period, and "Recursive" that it is updated every period. The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Term structure models				
Vasicek	-8.92	22.09	40.11	28.39
	(0.73)	(0.18)	(0.10)	(0.01)
CIR	-39.58	15.30	34.65	29.66
	(0.85)	(0.33)	(0.13)	(0.01)
$AFNS_0$	-5250.04	-11882.82	36.8	9.31
	(0.87)	(0.92)	(0.13)	(0.36)
$AFNS_0$ (uncorrected)	-471.70	10.58	39.78	-280.58
	(0.87)	(0.62)	(0.11)	(1.00)
$AFNS_3$	-14.36	21.75	46.72	28.69
	(0.72)	(0.29)	(0.09)	(0.01)
$AFNS_3$ (2nd moments)	-82.25	2.42	47.08	27.99
	(1.00)	(0.48)	(0.08)	(0.02)
PCA based forecasters				
PCA_3	-10.61	20.33	47.43	32.40
	(0.73)	(0.26)	(0.09)	(0.00)
PCA_4	-11.37	19.46	47.24	30.79
	(0.74)	(0.26)	(0.09)	(0.01)
PCA_5	-11.99	21.47	46.78	30.27
	(0.75)	(0.24)	(0.09)	(0.01)
PCA_6	-36.50	13.41	45.87	30.07
	(0.95)	(0.32)	(0.10)	(0.01)
Risk premium based forecasters				
Forward spreads	-20.56	13.57	47.80	33.52
	(0.89)	(0.34)	(0.08)	(0.00)
Yield spreads	-19.79	13.15	47.42	34.20
	(0.88)	(0.34)	(0.08)	(0.01)
Common factor based forecasters				
Initial	-10.68	23.07	43.77	36.57
	(0.76)	(0.21)	(0.12)	(0.00)
Recursive	-26.76	15.37	45.73	37.12
	(0.91)	(0.28)	(0.10)	(0.00)
Time series models				
HAR	3.46	-51.40	30.88	32.81
	(0.39)	(0.92)	(0.20)	(0.00)
Mean-reverting realized GARCH	-34.51	4.25	45.73	35.53
	(0.90)	(0.43)	(0.10)	(0.01)

Table 5: The incremental information on future volatility in the yield curve
This table shows results from regression Eq. (23), explaining yield volatility forecast errors from the specified models using three PCA factors fitted to past yield curves, based on Eq. (8). Reported values are F-statistics for joint tests of $\phi_1^{\tau,h} = 0$ (three coefficients). Asymptotic p-values in parentheses. The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	F-test			
	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Term structure models				
Vasicek	14.60	16.22	5.61	12.68
	(0.00)	(0.00)	(0.13)	(0.01)
CIR	39.75	32.14	3.92	24.43
	(0.00)	(0.00)	(0.27)	(0.00)
AFNS_0	6.88	6.67	9.37	0.92
	(0.08)	(0.08)	(0.02)	(0.82)
AFNS_3	23.91	53.53	6.62	14.04
	(0.00)	(0.00)	(0.09)	(0.00)
Time series models				
HAR	1.70	3.92	2.61	11.53
	(0.64)	(0.27)	(0.46)	(0.01)
Mean-reverting realized GARCH	8.48	16.30	4.91	15.10
	(0.04)	(0.00)	(0.18)	(0.00)

Table 6: Specification test for unspanned stochastic volatility factor

This table shows results from regression Eq. (24), explaining yield volatility forecast errors from the specified models augmented with three yield curve PCA factors as in Eq. (23) using a PCA factor fitted to past yield volatility forecast errors. Reported values are t-statistics for testing $\psi_1^{\tau,h} = 0$. Asymptotic p-values in parentheses, based on the Newey-West variance estimator with automatic lag selection of Andrews (1991). The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Vasicek	-1.22	-1.58	0.43	-2.25
	(0.22)	(0.11)	(0.67)	(0.02)
CIR	-2.13	-2.35	1.14	-2.70
	(0.03)	(0.02)	(0.25)	(0.01)
AFNS_0	-1.73	-2.06	0.39	-2.41
	(0.08)	(0.04)	(0.69)	(0.02)
AFNS_3	-2.13	-2.24	0.30	-3.09
	(0.03)	(0.03)	(0.76)	(0.00)

Table 7:
The incremental information on future yield volatility in historical volatility
This table shows results from the regression Eq. (23), using pre-averaged realized volatility from high-frequency data as predictor, to test for whether historical volatility contains incremental information about future volatility, relative to the specified models. Reported values are t-statistics for testing $\phi_1^{\tau,h} = 0$. Asymptotic p-values in parentheses, based on the Newey-West variance estimator with automatic lag selection of Andrews (1991). The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Term structure models				
Vasicek	0.32	-2.18	0.19	-0.21
	(0.75)	(0.03)	(0.85)	(0.84)
CIR	0.18	0.03	-1.47	-0.09
	(0.86)	(0.97)	(0.14)	(0.93)
$AFNS_0$	-1.56	-1.44	0.84	0.65
	(0.12)	(0.15)	(0.40)	(0.51)
$AFNS_3$	0.28	0.14	-0.38	-0.98
	(0.78)	(0.89)	(0.70)	(0.33)
$AFNS_3$ (2nd moments)	0.57	-0.20	-0.61	-0.10
	(0.57)	(0.84)	(0.54)	(0.92)
PCA based forecasters				
PCA_3	4.49	3.46	-0.69	0.70
	(0.00)	(0.00)	(0.49)	(0.48)
PCA_4	3.69	3.12	-0.72	0.78
	(0.00)	(0.00)	(0.47)	(0.44)
PCA_5	3.46	2.87	-0.77	-0.10
	(0.00)	(0.00)	(0.44)	(0.92)
PCA_6	0.08	0.45	-0.97	-0.80
	(0.94)	(0.65)	(0.33)	(0.43)
Risk premium based forecasters				
Forward spread	6.53	4.31	0.48	1.94
	(0.00)	(0.00)	(0.63)	(0.05)
Yield spread	6.61	4.60	0.85	2.58
	(0.00)	(0.00)	(0.40)	(0.01)
Common factor based forecasters				
Static	1.32	1.74	-0.77	-0.33
	(0.19)	(0.08)	(0.44)	(0.74)
Recursive	0.37	0.48	-0.60	-1.73
	(0.71)	(0.63)	(0.55)	(0.08)
Time series models				
HAR	-0.38	-0.16	-0.09	0.41
	(0.70)	(0.87)	(0.93)	(0.68)
Mean-reverting realized GARCH	4.42	5.47	1.06	1.95
	(0.00)	(0.00)	(0.29)	(0.05)

Table 8: Utility from yield volatility forecasts

This table presents average realized utility from using the specified yield volatility forecasters for portfolio allocation. In parentheses asymptotic p-values for a one-sided Diebold-Mariano test relative to a RW using the Newey-West variance estimator with automatic lag selection of Andrews (1991). A portfolio consisting of the risk-free asset generates utility zero. The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Term structure models				
Vasicek	0.75	1.08	-28.96	1.33
	(1.00)	(0.98)	(0.72)	(0.00)
CIR	1.24	1.28	-27.14	-5.01
	(0.93)	(0.69)	(0.69)	(0.83)
$AFNS_0$	1.08	0.95	-20.78	-0.99
	(1.00)	(1.00)	(0.57)	(0.79)
$AFNS_3$	0.99	1.11	-32.65	-6.34
	(1.00)	(0.99)	(0.69)	(0.86)
PCA based forecasters				
PCA_3	0.90	0.24	-97.02	1.08
	(1.00)	(1.00)	(1.00)	(0.03)
PCA_4	0.92	0.01	-96.59	1.11
	(1.00)	(1.00)	(1.00)	(0.02)
PCA_5	0.94	-0.03	-101.59	-0.41
	(1.00)	(0.99)	(1.00)	(0.75)
PCA_6	0.97	-0.01	-102.87	-0.19
	(1.00)	(0.99)	(1.00)	(0.70)
Risk premium based forecasters				
Forward Spread	1.01	0.92	0.02	1.25
	(1.00)	(1.00)	(0.00)	(0.01)
Yield Spread	1.01	0.92	0.98	1.36
	(1.00)	(1.00)	(0.00)	(0.00)
Common factors based forecasters				
Static	1.13	0.86	-26.43	0.99
	(1.00)	(0.99)	(0.77)	(0.06)
Recursively	0.97	-0.02	-77.71	-0.22
	(1.00)	(0.99)	(0.99)	(0.72)
Term structure models including second moments				
CIR	0.52	0.63	0.59	0.49
	(1.00)	(1.00)	(0.00)	(0.38)
$AFNS_3$	0.99	1.02	-34.64	-7.98
	(1.00)	(1.00)	(0.70)	(0.92)
Time series models				
HAR	1.29	1.01	1.00	1.36
	(0.95)	(1.00)	(0.00)	(0.00)
Mean-reverting realized GARCH	0.92	0.05	1.03	1.37
	(1.00)	(0.96)	(0.00)	(0.00)
RW	1.35	1.31	-17.95	0.38
	(-)	(-)	(-)	(-)

Table 9: Yield volatility forecasts using nonlinearities in yield curve factors
This table displays R^2_{OoS} measures in percent relative to a RW for yield volatility forecasts allowing linear and nonlinear terms as well as interactions in the three leading yield curve factors extracted by PCA in Eq. (8). The optimal combination of terms is selected by BIC. The label "initial" indicates that the selection is based on the initial estimation period, and "recursive" that it is updated every period. In parentheses asymptotic p-values for a one-sided Diebold-Mariano test using the Newey-West variance estimator with automatic lag selection of Andrews (1991). The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Maturity-specific, initial	-36.51	11.57	-0.65	18.52
	(0.95)	(0.35)	(0.51)	(0.16)
Maturity-specific, recursive	6.08	30.28	-29.41	-16.64
	(0.35)	(0.16)	(0.67)	(0.64)
Common factor, initial	5.74	31.39	45.89	39.30
	(0.36)	(0.14)	(0.08)	(0.00)
Common factor, recursive	-14.76	23.24	35.96	33.76
	(0.78)	(0.20)	(0.14)	(0.00)

Table 10: Trade-off between explaining yields and forecasting volatility
This table displays R^2_{OoS} measures in percent relative to a RW for all combinations of three of the first six PCA yield factors from Eq. (8). In parentheses asymptotic p-values for a one-sided Diebold-Mariano test using the Newey-West variance estimator with automatic lag selection of Andrews (1991). The initial estimation period ranges from January 2, 2000 to June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
PCA_{456}	-11.93	20.67	46.01	36.30
	(0.75)	(0.25)	(0.10)	(0.00)
PCA_{356}	-5.13	23.96	48.86	41.47
	(0.63)	(0.22)	(0.09)	(0.00)
PCA_{346}	-12.99	19.72	49.05	39.00
	(0.81)	(0.25)	(0.09)	(0.00)
PCA_{345}	-5.91	22.52	49.10	39.84
	(0.64)	(0.22)	(0.09)	(0.00)
PCA_{256}	-7.55	26.79	48.39	39.60
	(0.66)	(0.19)	(0.09)	(0.00)
PCA_{246}	-8.82	22.86	48.40	34.54
	(0.69)	(0.23)	(0.09)	(0.00)
PCA_{245}	-13.01	23.72	47.86	35.27
	(0.78)	(0.20)	(0.09)	(0.00)
PCA_{236}	-4.58	24.52	48.20	31.67
	(0.61)	(0.21)	(0.10)	(0.01)
PCA_{235}	-6.40	25.37	48.17	31.16
	(0.64)	(0.18)	(0.09)	(0.01)
PCA_{234}	-6.11	22.81	48.47	28.38
	(0.64)	(0.21)	(0.09)	(0.01)
PCA_{156}	-15.96	15.31	47.37	34.30
	(0.80)	(0.32)	(0.08)	(0.01)
PCA_{146}	-19.43	10.95	47.43	30.16
	(0.86)	(0.37)	(0.08)	(0.02)
PCA_{145}	-16.75	17.11	46.95	32.99
	(0.83)	(0.29)	(0.08)	(0.01)
PCA_{136}	-15.81	13.94	48.54	37.39
	(0.84)	(0.33)	(0.09)	(0.00)
PCA_{135}	-9.46	20.50	48.43	39.66
	(0.71)	(0.26)	(0.09)	(0.00)
PCA_{134}	-10.56	17.66	48.75	37.65
	(0.73)	(0.29)	(0.08)	(0.00)
PCA_{126}	-14.39	17.05	47.90	37.20
	(0.77)	(0.30)	(0.08)	(0.00)
PCA_{125}	-16.86	21.46	47.73	37.90
	(0.83)	(0.24)	(0.08)	(0.00)
PCA_{124}	-15.44	18.51	48.01	34.07
	(0.81)	(0.27)	(0.08)	(0.01)
PCA_{123}	-10.61	20.33	47.43	32.40
	(0.63)	(0.25)	(0.08)	(0.01)

Table 11: Yield volatility forecasts from cross section of coupon bond prices
This table displays R^2_{OoS} measures in percent relative to a RW for the CIR model estimated on a cross section of coupon bonds. $\text{CIR}_{\mathbb{Q}}$ indicates that forecasting is based on cross-sectionally estimated $\mathbb{Q}$-parameters and short rates, and $\text{CIR}_{\mathbb{P}}$ that the market price of risk is estimated based on the $\mathbb{Q}$-parameters, the ten most recent fitted short rates, and Euler discretization. In parentheses asymptotic p-values for a one-sided Diebold-Mariano test using the Newey-West variance estimator with automatic lag selection of Andrews (1991). The initial estimation period ranges from January 2, 2000 to June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
$\text{CIR}_{\mathbb{Q}}$	-48.91	8.46	47.24	27.35
	(0.91)	(0.43)	(0.10)	(0.04)
$\text{CIR}_{\mathbb{P}}$	-55.02	4.97	46.81	32.87
	(0.93)	(0.46)	(0.10)	(0.03)

Figure 1: High-frequency yields

This figure shows the time series of high-frequency annualized yields in percent, by maturity, along with a three-dimensional view of the evolution through time of the yield curves. The sample spans the period from January 2, 2000, through October 31, 2016.

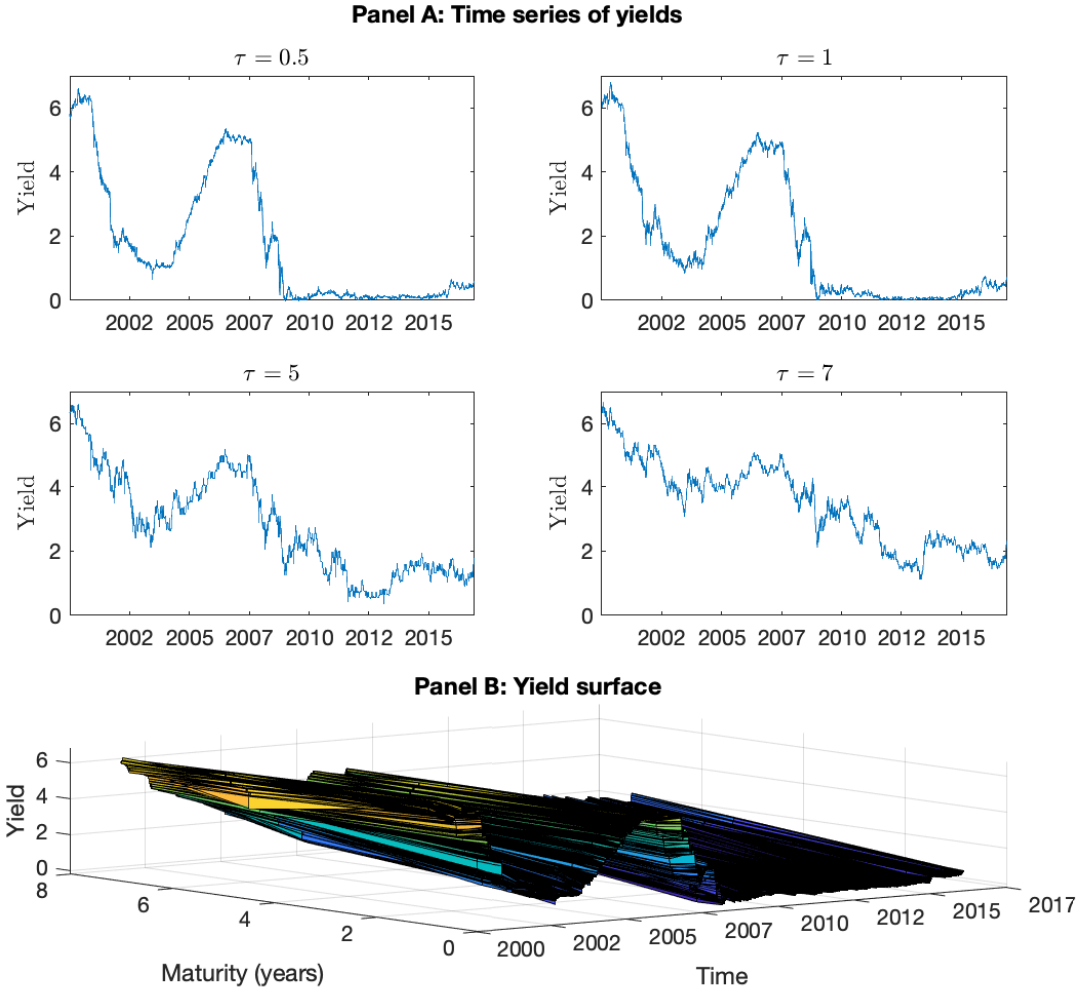


Figure 2: Annualized pre-averaged realized yield volatility

This figure shows the time series of daily annualized pre-averaged realized one-month yield volatilities in percent, by maturity, along with a three-dimensional view of the evolution through time of the term structure of volatilities (log variances). The sample spans the period from January 2, 2000, through October 31, 2016.

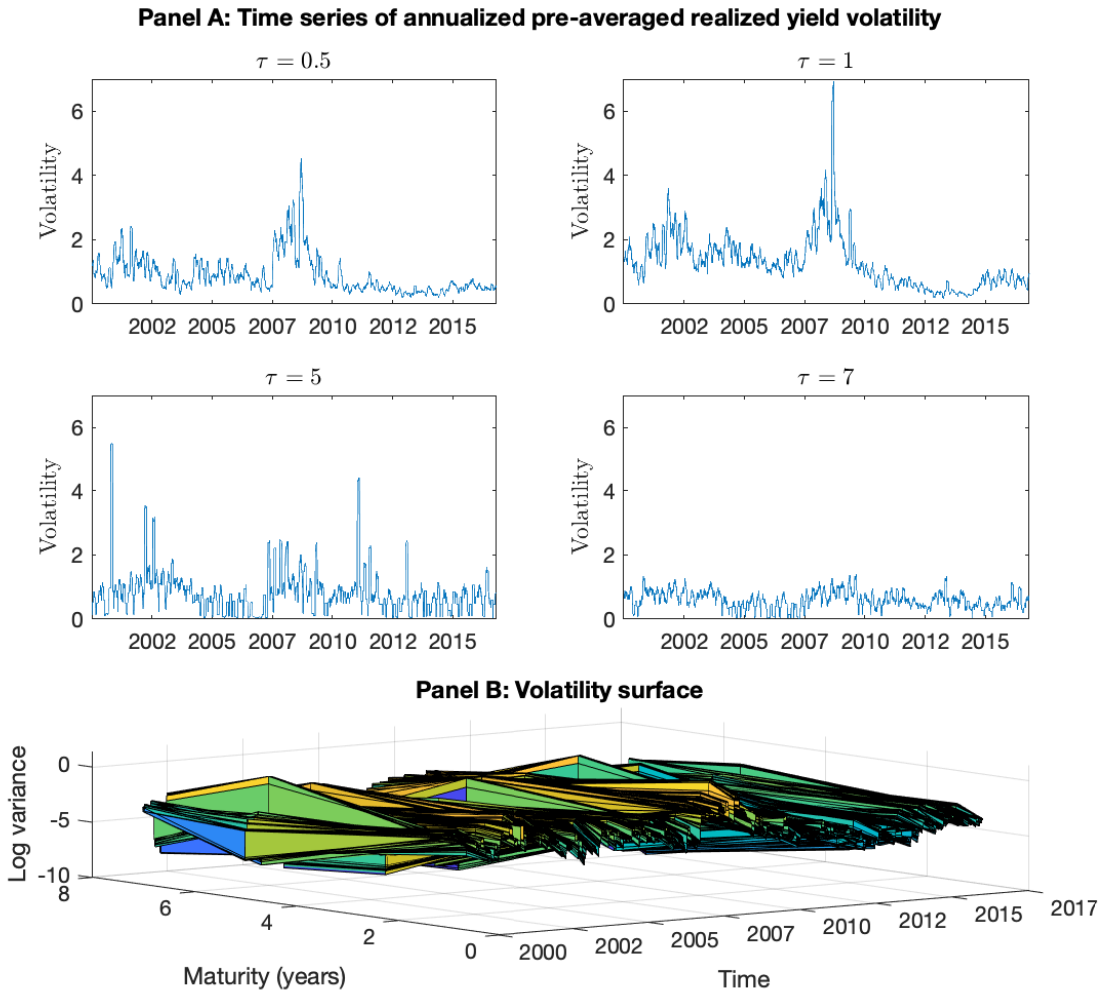
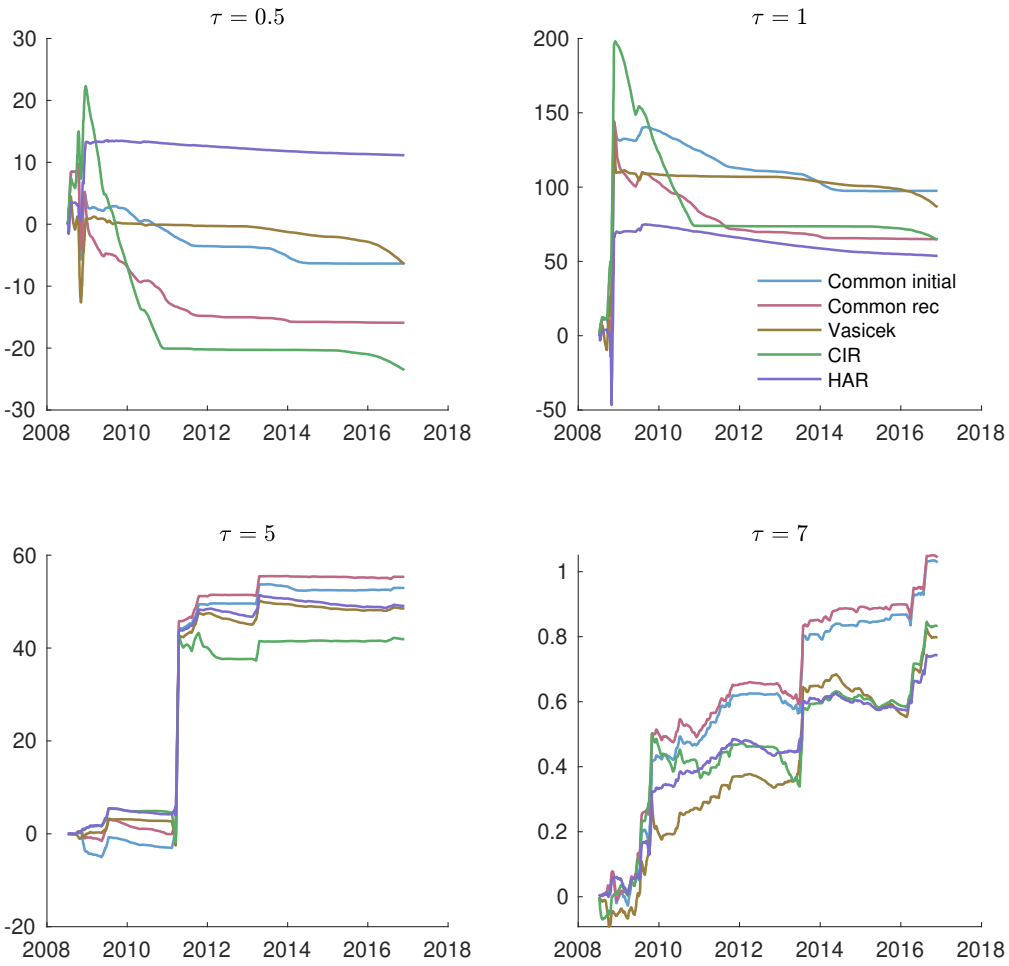


Figure 3: Yield volatility forecast errors relative to RW

This figure shows yield volatility forecast errors over time, by maturity, for selected forecasting methods. Cumulative squared errors for a given forecasting method are subtracted from those for the RW, so that an increasing curve indicates better forecasting than by RW, and vice versa. The sample spans the period from January 2, 2000, through October 31, 2016.



A. From futures prices to yield curves

A.1. From futures prices to coupon bond prices

The seller (the short) has two options included in the Treasury futures. The first regards which bond to deliver, and the second the delivery date within the delivery month. For the first option, the underlying assets of the Treasury futures (see Section 3) are hypothetical bonds with a notional yield of 6% throughout our sample period. The seller must deliver an actual bond, selected from a delivery basket constructed for each futures contract according to CME requirements, with a conversion factor for each bond. The conversion factor converts the bond price into "...the approximately decimal price at which \$1 par of the security would trade as if it had a 6% yield-to-maturity."[17] The formula is

$$
\begin{aligned}
f &= a\left(\frac{coupon}{2} + c + d\right) - b \qquad \text{(A.1)}\\
a &= \left(\frac{1}{1.03}\right)^{\frac{v}{6}}\\
b &= \frac{coupon}{2}\frac{6-v}{6}\\
c &= \begin{cases} \left(\frac{1}{1.03}\right)^{2n} & \text{if } z < 7 \\ \left(\frac{1}{1.03}\right)^{2n+1} & \text{otherwise} \end{cases}\\
d &= \frac{coupon}{0.06}(1-c)\\
v &= \begin{cases} z & \text{if } z < 7 \\ 3 & \text{if } z \geq 7 \text{ for 10-year note, long term bond} \\ z-6 & \text{if } z \geq 7 \text{ for 5-year note} \end{cases}
\end{aligned}
$$

with *coupon* the annualized coupon, n the number of (whole) years from the first day of the delivery month to maturity, and z the number of months between n and the maturity date, rounded to nearest quarter for the 10-year note and the long term Treasury bond, and nearest month for the 5-year note. The delivery bond (or note) must have a maturity between 4.17 and 5.25 years for the 5-year futures, between 6.5 and 10 years for the 10-year futures, and between 15 and 25 years for the long term Treasury bond futures.

[17] https://www.cmegroup.com/trading/interest-rates/calculating-us-treasury-futures-conversion-factors.html

Given the conversion factor f_t from Eq. (A.1), the invoice amount is

$$I_t = f_t \cdot F_t + a_t \,, \tag{A.2}$$

with F_t the futures price, and a_t accrued interest. On delivery, the seller pays the basis

$$\pi_t = S_t - I_t \,, \tag{A.3}$$

with S_t the cash bond price (including accrued interest), and therefore selects the cheapest-to-deliver (CTD) bond fulfilling the futures contract specifications and minimizing Eq. (A.3) (maximizing the implied repo rate).

The second option regards the delivery date. For the purpose of backing out the futures-implied bond price, we assume that both the CTD bond and delivery date are known. We set the delivery date to the first working day of the delivery month, and use the CRSP data on Treasuries to construct the delivery basket. The price of the futures is then given as

$$I_t = (S_t - C_t)e^{r_t T} \,, \tag{A.4}$$

with S_t the spot price of the CTD bond (including accrued interest), C_t the present value of coupon payments before delivery, T the time to delivery, and r_t the risk-free rate. For the latter, we use the 3-month yield from the daily panel, $r_t = y_t^{1/4}$ (Section 3.4). The 3-month rate should be more market-based than, say, a 1-month rate, and a good short rate proxy, cf. Chapman et al. (1999).

In brief, the recipe for constructing the futures-implied bond price S_t is as follows:

(i) Construct the delivery basket from the CRSP data on the first trading day of the delivery month.

(ii) Find the cheapest-to-deliver bond minimizing the basis π_t from Eq. (A.3) within the delivery basket from (i).

(iii) Use Eq. (A.2) to calculate the invoice amount I_t for the CTD bond from (ii).

(iv) Substitute the invoice amount from (iii) for I_t in Eq. (A.4), and back out the futures-implied bond price by isolating S_t.

A.2. From coupon bond prices to yield curves

A variety of methods is available for extracting the yield curve from the cross section of coupon bond prices. In the literature on yield curves at high frequency, some version of cubic spline is usually employed, e.g., Andersen and Benzoni (2010) and Cieslak and Povala (2016), due to the small number of observations in the cross section, and the need for flexibility to fit the curve reasonably. As a consequence, interpolation between points of principal payments is difficult. Following Andersen and Benzoni (2010), we use the method of Waggoner (1997). This assumes that the discount function $B_t^\tau = \exp(-\tau y_t^\tau)$ is determined by a cubic spline $h(\cdot, \Psi)$, i.e., $\tau y_t^\tau = h(\tau, \Psi)$, with Ψ a set of parameters (suppressing dependence on t). The yield curve is split along the maturity axis using K knot points, $0 < \tau_1 < \ldots < \tau_K$, with τ_K the maximum maturity of the bonds considered. The cubic spline is restricted such that $h(\cdot, \Psi)$ and its two first derivatives are continuous. This implies that one parameter is added for each additional knot point. Fisher et al. (1995) and Dai et al. (2007) set K to approximately one third of the number of bonds included, while Andersen and Benzoni (2010) and Cieslak and Povala (2016) set K equal to the number of bonds. We follow the high-frequency literature and set the number of knot points equal to the number of bonds on which we calibrate the yield curve, i.e., $K = 5$. We let $\{s_k\}_{k=1}^K$ denote the set of knot points.

To calibrate $h(\tau, \Psi)$, we follow Fisher et al. (1995) and use the simple parametrization of the cubic B-spline basis. Any cubic spline can be constructed as a linear combination of B-splines, $h(\tau, \Psi) = \phi(\tau)\Psi$, where Ψ is a $\kappa \times 1$ vector of coefficients, and $\phi(\cdot)$ is a cubic B-spline basis, i.e., a vector of $\kappa = K + 2$ cubic B-splines satisfying

$$\phi_k^r(\tau) = \phi_k^{r-1}(\tau)\frac{\tau - d_k}{d_{k+r-1} - d_k} - \phi_{k+1}^{r-1}(\tau)\frac{d_{k+r} - \tau}{d_{k+r} - d_{k+1}}\,.$$

Here, $r = 4$ for a cubic spline, and d_k is an augmented set of knot points, $d_1 = d_2 = d_3 = s_1$, $d_{k+3} = s_k$ for $1 \leq k \leq \kappa$, and $d_{\kappa+4} = d_{\kappa+5} = d_{\kappa+6} = s_\kappa$.

The price of a coupon bond is written as

$$P(\Psi) = \sum_{j=1}^{J} C_j \exp(-h(t_j, \Psi)) \,,$$

with J the number of payments (coupons and principal) over the remaining lifetime of the bond, and C_j the j^{th} payment, due t_j periods hence. The yield curve is estimated using penalized nonlinear least squares (PNLS),

$$\hat{\Psi} = \arg\min_{\Psi} \sum_{i=1}^{K} (P_i - \hat{P}_i(\Psi))^2 + \int_0^{\tau_K} \lambda(s) h''(s)^2 ds \,,$$

with the penalty term $\lambda(s)$ calibrated to be more severe at longer maturities, following Waggoner (1997), Dai et al. (2007), and Andersen and Benzoni (2010). This implies that the yield curve is flexible at the short end, and reduces oscillations at the long end. We find that this procedure results in a better fit to bond prices than the Nelson and Siegel (1987) approach used in Faust et al. (2007).

Upon estimation, four points are read off the resulting high-frequency (1 minute) yield curves, at $\tau = 0.5$, 1, 5, and 7 years.

B. Liquidity of Treasury futures

Table A.2 shows the percentage of high-frequency intervals containing a Treasury futures price observation, and hence affording unique yield identification, at the 1, 5, and 10 minute sampling frequencies, by maturity. All three maturities, 5 and 10 years, and long term, are included in the table, because all three Treasury futures are used in curve fitting, although only implied yields at maturities up to $\tau = 7$ are used in the subsequent analysis (see Section 3.2 and Appendix A). From the table, around 83% to 95% of the 1 minute intervals contain an observation. This increases to around 97% and 99% for 5 and 10 minute intervals, respectively. The lower panel of the table shows the corresponding percentages when discarding trading days with no trading activity for more than an hour. All numbers increase by around 0.5-1.0%. This suggests that the discarded days do not contain many uniquely identified observations and, hence, it makes sense to discard them.

Figure A.1 shows the evolution over time in the daily percentage of intervals containing a futures price observation. There is a jump between 2003 and 2004, due to the introduction of the electronic trading pit in 2004, and clearly fewer intervals containing futures price observations at the 1 minute frequency than at 5 and 10 minutes, at least until 2008. This suggests that the high-frequency Treasury futures data are likely afflicted with market microstructure noise.

INSERT TABLE A.2 AND FIGURE A.1 ABOUT HERE

C. Models

C.1. The Vasicek model

The model of Vasicek (1977) is an $A_0(1)$ model in which the dynamics of the short rate $r_t = y_t^0$ are given by

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \,.$$

We assume that the market price of risk is completely affine, i.e., $\lambda_t = \lambda$. Thus,

$$\tilde{\theta} = \theta - \frac{\sigma\lambda}{\kappa}, \;\; \tilde{\kappa} = \kappa \,.$$

The solution to the Ricatti equations (6) is given by

$$\begin{aligned} B(\tau) &= \frac{1}{\tilde{\kappa}}(1 - e^{-\tilde{\kappa}\tau}) \,, \\ A(\tau) &= \left(\tilde{\theta} - \frac{\sigma^2}{2\tilde{\kappa}^2}\right)(\tau - B(\tau)) + \frac{\sigma^2}{4\tilde{\kappa}}B(\tau)^2 \,. \end{aligned}$$

The conditional variance of the future short rate is

$$Var_t(r_{t+h}) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa h}) \,.$$

C.2. The Cox-Ingersoll-Ross Model

The model of Cox et al. (1985) is an $A_1(1)$ model with

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t\,.$$

We adopt the completely affine market price of risk specification, $\lambda_t = \lambda\sqrt{r_t}/\sigma$. Thus,

$$\tilde{\kappa} = \kappa + \lambda,\ \tilde{\theta} = \frac{\kappa\theta}{\kappa + \lambda}\,.$$

The solution to the Ricatti equations (6) is given by

$$\begin{aligned} B(\tau) &= \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \tilde{\kappa})(e^{\gamma\tau} - 1) + 2\gamma}\,, \\ A(\tau) &= -\frac{2\tilde{\kappa}\tilde{\theta}}{\sigma^2}\left[\log(2\gamma) + \frac{1}{2}(\tilde{\kappa} + \gamma) - \log\left((\gamma + \tilde{\kappa})(e^{\gamma\tau} - 1) + 2\gamma\right)\right]\,, \end{aligned}$$

with γ given by

$$\gamma = \sqrt{\tilde{\kappa}^2 + 2\sigma^2}\,.$$

The conditional variance of the short rate is

$$Var_t(r_{t+h}) = \frac{\sigma^2 r_t}{\kappa}(e^{-\kappa h} - e^{-2\kappa h}) + \frac{\sigma^2\theta}{2\kappa}(1 - e^{-\kappa h})\,.$$

C.3. The arbitrage-free Nelson-Siegel Model with deterministic volatility

The AFNS_0 model of Christensen et al. (2011) is an $A_0(3)$ model. The real-world dynamics of the state variables are given by

$$dX_t = \kappa(\theta - X_t)dt + \Sigma dW_t\,,$$

where X_t and θ are 3×1 vectors, and κ and Σ are 3×3 matrices. The model is constructed to make the shape of the yield curve resemble the Nelson and Siegel (1987) parametrization

by assuming that the market price of risk is essentially affine and restricting

$$\tilde{\kappa} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & \lambda & \lambda \end{pmatrix},$$

$$\tilde{\theta} = 0_{3\times 1}\,.$$

The yield curve is then

$$y_t^\tau = X_t^1 + \frac{1-e^{-\lambda\tau}}{\lambda\tau} X_t^2 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) X_t^3 - \frac{A(\tau)}{\tau}\,. \tag{C.5}$$

This corresponds to the Nelson and Siegel (1987) shape, except for the term $\frac{A(\tau)}{\tau}$, for which the closed form is given in Christensen et al. (2011).

C.4. The arbitrage-free Nelson-Siegel model with stochastic volatility

The AFNS$_3$ stochastic (or state-dependent) volatility model of Christensen et al. (2010) is an $A_3(3)$ model. Under $\mathbb{Q}$, the dynamics of the state variables are restricted to

$$\begin{aligned} dX_t = &\begin{pmatrix} \varepsilon & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \left[\begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \tilde{\theta}_3 \end{pmatrix} - X_t \right] dt \\ &+ \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ 0 & \sigma_{2,2} & 0 \\ 0 & 0 & \sigma_{3,3} \end{pmatrix} \begin{pmatrix} \sqrt{X_{1,t}} & 0 & 0 \\ 0 & \sqrt{X_{2,t}} & 0 \\ 0 & 0 & \sqrt{X_{3,t}} \end{pmatrix} dW_t^{\mathbb{Q}}\,. \end{aligned}$$

The Ricatti equations (6) are then given by

$$\begin{aligned} dB_1(\tau) &= 1 - \varepsilon B_1(\tau) - \frac{1}{2}\sigma_{1,1}^2\,, \\ dB_2(\tau) &= 1 - \lambda B_2(\tau) - \frac{1}{2}\sigma_{2,2}^2\,, \\ dB_3(\tau) &= \lambda B_2(\tau) - \lambda B_3(\tau) - \frac{1}{2}\sigma_{3,3}^2\,, \\ dA(\tau) &= B(\tau)'\tilde{\kappa}\tilde{\theta}\,. \end{aligned}$$

We solve the ODEs numerically for each trial parameter vector in the iterative estimation procedure. For $\varepsilon \to 0$, the yield curve converges to Nelson and Siegel (1987) shape, except for the modification by $\frac{A(\tau)}{\tau}$, as in Eq. (C.5). Following Christensen et al. (2010), we set $\varepsilon = 10^{-6}$, restrict κ to be diagonal, and adopt the extended affine specification for the market price of risk. To ensure that the factors stay positive, we impose the Feller condition under both probability measures.

D. Estimation Procedure

The term structure models are estimated using the Kalman filter, following Duffee (2002), Christensen et al. (2011), and others. Let y_t, $t = 1, \ldots, T$, be the N-vector of observed yields, with t counting the time increment between observations of $\Delta = 1/250$ years, $\tilde{A} = (\frac{A(\tau_1)}{\tau_1}, \ldots, \frac{A(\tau_N)}{\tau_N})'$, and $\tilde{B}$ the $d \times N$ matrix with columns $\frac{B(\tau_i)}{\tau_i}$. By Eqs. (18) and (19), the state space system is thus given by the measurement and transition equations

$$y_t = \tilde{A} + \tilde{B}' X_t + \varepsilon_t \,, \tag{D.6}$$

$$X_t = C_\Delta + D'_\Delta X_{t-1} + \eta_t \,, \tag{D.7}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_\Delta & 0 \\ 0 & Q_{t,\Delta} \end{pmatrix} \right] .$$

The conditional expectation and variance of the state process are given by

$$\mathbb{E}_{t-1}(X_t) = (I - e^{-\kappa\Delta})\theta + e^{-\kappa\Delta} X_{t-1} = C_\Delta + D'_\Delta X_{t-1} \,, \tag{D.8}$$

$$\begin{aligned} Var_{t-1}(X_t) = \int_{t-1}^{t} & \exp(-\kappa(t-u))\sigma(\mathbb{E}_{t-1}(X_u))\sigma(\mathbb{E}_{t-1}(X_u))' \\ & \exp(-\kappa(t-u))' du = Q_{t,\Delta} \,, \end{aligned} \tag{D.9}$$

where $\exp(\cdot)$ is the matrix exponential, and $\sigma(\cdot) = \Sigma\sqrt{S(\cdot)}$, with $S(\cdot)$ from Eq. (3). For example, $\sigma(X_u) = \text{diag}(\sigma_{i,i}\sqrt{X_{i,u}})$ in the model in Appendix C.4. By Eq. (D.8), the transition parameters in Eqs. (19) and (D.7) are $C_\Delta = (I_d - \exp(-\kappa\Delta))\theta$ and $D_\Delta =$

$\exp(-\kappa\Delta)$. Write $Y_t = (y_1, \ldots, y_t)$ for the data through t and define $X_{t|t-1} = \mathbb{E}(X_t \mid Y_{t-1})$, $\Sigma_{t|t-1} = Var(X_t \mid Y_{t-1})$, $X_{t|t} = \mathbb{E}(X_t \mid Y_t)$, and $\Sigma_{t|t} = Var(X_t \mid Y_t)$. Suppressing Δ for notational ease, the prediction step is then

$$X_{t|t-1} = C + D' X_{t-1|t-1}\,, \tag{D.10}$$

$$\Sigma_{t|t-1} = D'\Sigma_{t-1|t-1} D + Q_t\,, \tag{D.11}$$

where Q_t is the conditional variance of the state variables from Eq. (D.9), while $D'\Sigma_{t-1|t-1}D$ reflects conditional mean variation, cf. footnote 3. The update step is

$$X_{t|t} = X_{t|t-1} + \Sigma_{t|t-1}\tilde{B} F_t^{-1} v_t\,, \tag{D.12}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\tilde{B} F_t^{-1} \tilde{B}'\Sigma_{t|t-1}\,, \tag{D.13}$$

with the one step ahead prediction error and its variance given by

$$v_t = y_t - \tilde{A} - \tilde{B}' X_{t|t-1}\,, \tag{D.14}$$

$$F_t = Var(v_t) = \tilde{B}'\Sigma_{t|t-1}\tilde{B} + H\,. \tag{D.15}$$

The filter is initiated by setting $X_{0|0}$ and $\Sigma_{0|0}$ equal to the unconditional expected value and variance. From Eq. (2), $X_{0|0} = \theta$, and from Eq. (D.7), $\Sigma_{0|0} = D'\Sigma_{0|0}D + Q$, with Q from Eq. (D.9) at $\mathbb{E}_{t-1}(X_u) = \theta$. Thus, $vec(\Sigma_{0|0}) = (I_{d^2} - D' \otimes D')^{-1}\, vec(Q)$.

Given the Gaussianity assumptions on measurement and transition equations, the parameters ψ are estimated by maximizing the prediction error decomposition of the conditional log likelihood function

$$L(\psi) = \sum_{t=1}^{T} (-\frac{1}{2} N \log(2\pi) - \frac{1}{2}\log(\det F_t) - \frac{1}{2} v_t' F_t^{-1} v_t)\,. \tag{D.16}$$

The states X_t are Gaussian for the Vasicek and AFNS_0 models. For the square root processes (the CIR and AFNS_3 models), estimation based on Eq. (D.16) amounts to QML. Here, the Gaussian approximation implies that states are not restricted to be positive. In

this case, following Chen and Scott (2003), we truncate states at 0. Given the large number of parameters in some of the models, we use the global optimizer differential evolution with several starting values. For every 250 observations, we reestimate parameters using differential evolution. In intermediate periods we use local optimizers.

D.1. Predictive regression correction

From Eq. (D.6), $var(y_t \mid X_{t-1}) = \tilde{B}' Var_{t-1}(X_t)\tilde{B} + H$, suppressing dependence on the time increment Δ. Using this and Eqs. (D.9), (D.11) and (D.15),

$$\begin{aligned} Var(y_t \mid Y_{t-1}) &= Var(v_t) \\ &= \tilde{B}'(D'\Sigma_{t-1|t-1}D + Var_{t-1}(X_t))\tilde{B} + H \\ &= \tilde{B}' D'\Sigma_{t-1|t-1}D\tilde{B} + Var(y_t \mid X_{t-1})\,. \end{aligned} \tag{D.17}$$

Thus, the conditional variance of yields is given by that corresponding to perfect observation of state variables through $t-1$, i.e., the second term in Eq. (D.17), with an adjustment for conditional mean variation due to imperfect state observation given by the first term in Eq. (D.17) (see footnote 3). When forecasting using the predictive regression in Eq. (1), with Z_t given by Eq. (7), then Z_t corresponds to the second term in Eq. (D.17). Forecasting therefore involves three modifications. First, the filtered state $X_{t|t}$ is used in place of X_t in Eq. (7). Second, using Eq. (1), $\alpha^{\tau,h}$ and $\beta^{\tau,h}$ provide empirical predictive regression corrections for the bias stemming from conditional mean variation, i.e., the first term in Eq. (D.17). Finally, the estimated coefficents in Eq. (1) reflect the history of realized volatilites, which are part of investor's information set, thereby conditioning on a larger information set than that based on daily yields in the Kalman filter, and providing a computationally simple alternative to the filter including second moments below.

D.2. Conditional moments of state process

By the results in Fackler (2000), also used in Jacobs and Karoui (2009), the first two conditional moments can be written as

$$\mathbb{E}_t(X_{t+h}) = C_h + D_h' X_t \,, \tag{D.18}$$

$$vec(Var_t(X_{t+h})) = \zeta_{0,h} + \zeta_{1,h}' X_t \,, \tag{D.19}$$

with $C_h = (I_d - \exp(-\kappa h))\theta$ and $D_h = \exp(-\kappa h)$, see Eqs. (D.8)-(D.9). The $d^2 \times 1$ vector of intercepts $\zeta_{0,h}$ in Eq. (D.19) is determined by

$$\begin{pmatrix} C_h \\ \zeta_{0,h} \end{pmatrix} = (I_{d+d^2} - \exp(\Pi h))\Pi^{-1}\Theta \,, \tag{D.20}$$

and the $d \times d^2$ matrix of slope terms $\zeta_{1,h}$ by

$$\begin{pmatrix} D_h \\ \zeta_{1,h}' \end{pmatrix} = \exp(-\Pi h) \begin{bmatrix} I_d \\ 0 \end{bmatrix} . \tag{D.21}$$

In Eq. (D.20), the $(d + d^2)$-vector Θ is

$$\Theta = \begin{bmatrix} \kappa\theta \\ (\Sigma \otimes \Sigma)\mathcal{D}\alpha \end{bmatrix} ,$$

with $\alpha = 0$ in Eq. (3) for the models we consider, and $\mathcal{D}$ given by

$$\mathcal{D}_{i,j} = \begin{cases} 1 & \text{if } i = (j-1)d + j \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

Further,

$$\Pi = \begin{bmatrix} \kappa & 0 \\ -(\Sigma \otimes \Sigma)\mathcal{D}\mathcal{B} & \kappa \otimes I_d + I_d \otimes \kappa \end{bmatrix} ,$$

a $(d + d^2) \times (d + d^2)$ matrix, with the i^{th} row of $\mathcal{B}$ given by β_i' from Eq. (3), and $\operatorname{rank} \mathcal{B} = m \leq d$.

D.3. Conditional moments of yields

By Eqs. (5) and (D.18), the conditional means of the yields are

$$\mathbb{E}_t(y_{t+h}) = \tilde{A} + \tilde{B}' \left(C_h + D_h' X_t\right) . \tag{D.22}$$

Using Eqs. (7) and (D.19), the conditional yield variances are

$$\begin{aligned} Var_t(y^\tau_{t+h}) &= \frac{1}{\tau^2} B(\tau)' Var_t(X_{t+h}) B(\tau) \\ &= \frac{1}{\tau^2} vec(B(\tau) \otimes B(\tau))' Var_t(X_{t+h}) \\ &= b_0^{\tau,h} + b_1^{\tau,h'} X_t, \qquad \text{(D.23)} \\ b_0^{\tau,h} &= \frac{1}{\tau^2} vec(B(\tau) \otimes B(\tau))' \zeta_{0,h} , \qquad \text{(D.24)} \\ b_1^{\tau,h\prime} &= \frac{1}{\tau^2} vec(B(\tau) \otimes B(\tau))' \zeta_{1,h}' , \qquad \text{(D.25)} \end{aligned}$$

where $b_0^{\tau,h}$ is a scalar, and $b_1^{\tau,h}$ a $d \times 1$ vector.

D.4. Including second moments in the estimation

The measurement equation Eq. (D.6) can be extended to include realized volatility,

$$\begin{pmatrix} y_{t+h} \\ V_{t+h} \end{pmatrix} = \begin{pmatrix} \tilde{A} + \tilde{B}' X_{t+h} \\ Var_t(y_{t+h}) \end{pmatrix} + \begin{pmatrix} H & 0 \\ 0 & H^V \end{pmatrix} \tilde{\varepsilon}_{t+h} , \tag{D.26}$$

where V_t is stacking the N_V realized yield volatilities V_t^τ for which high-frequency data are available ($N_V = 4$ in our implementation), the error term is stacking ε_{t+h} from Eq. (D.6) and $\tilde{u}^{\tau,h}_{t+h}$ from Eq. (20), $\tilde{\varepsilon}_{t+h} \sim \mathcal{N}(0, I_{N+N_V})$, and $H^V = H_h^V$, an $N_V \times N_V$ matrix with $\sigma^2_{\tau,h}$ along the diagonal. The prediction step is then

$$\begin{aligned} y_{t|t-1} &= \tilde{A} + \tilde{B}' X_{t|t-1} , \qquad \text{(D.27)} \\ V_{t|t-1} &= b_0 + b_1' X_{t|t-1} + H^V , \qquad \text{(D.28)} \end{aligned}$$

with $X_{t|t-1}$ from Eq. (D.10), $H^V = H_1^V$, b_0 an N_V-vector stacking $b_0^{\tau,h}$ from Eq. (D.24), b_1' an $N_V \times d$ matrix stacking $b_1^{\tau,h\prime}$ from Eq. (D.25), and $h = 1$ in all of these, for estimation in daily data. The update step is

$$X_{t|t} = X_{t|t-1} + \Sigma_{t|t-1} \begin{bmatrix} \tilde{B} \\ b_1 \end{bmatrix}' F_t^{-1} v_t \,, \tag{D.29}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} \begin{bmatrix} \tilde{B} \\ b_1 \end{bmatrix}' F_t^{-1} \begin{bmatrix} \tilde{B} \\ b_1 \end{bmatrix} \Sigma_{t|t-1} \,, \tag{D.30}$$

where v_t is the one step ahead prediction error or innovation

$$v_t = \begin{pmatrix} y_t - \tilde{A} - \tilde{B}' X_{t|t-1}, \\ V_t - b_0 - b_1' X_{t|t-1} - H^V \end{pmatrix} ,$$

and F_t the prediction error variance,

$$F_t = \begin{pmatrix} \tilde{B}\Sigma_{t|t-1}\tilde{B}' + H & b_1\Sigma_{t|t-1}\tilde{B}' \\ \tilde{B}\Sigma_{t|t-1}b_1' & b_1\Sigma_{t|t-1}b_1' + H^V \end{pmatrix} .$$

Because Eq. (D.26) uses conditional volatility rather than integrated volatility for the state variable, cf. footnote 3, the term involving H^V actually captures both the error in realized relative to integrated volatility and conditional mean variation corresponding to $D'\Sigma_{t-1|t-1}D$ in Eqs. (D.11) and (D.17). We find in our empirical work that adjustment using the recursive regressions in Eq. (1) performs better for forecasting purposes than the second moment extension of the Kalman filter.

Table A.1:
Correlation between futures-implied and Gürkaynak et al. (2007) yields
This table shows the sample correlation between end-of-day futures-implied yields and yields constructed using Gürkaynak et al. (2007) parameter estimates. The sample spans the period from January 2, 2000, through October 30, 2016.

$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
99.84%	99.86%	99.76%	99.58%

Table A.2: Liquidity of Treasury futures

This table shows the fraction of high-frequency intervals containing a Treasury futures price observation, by underlying maturity and sampling frequency. The full sample is included in the upper panel. In the lower panel, days with no trading activity for more than one hour are discarded. The sample spans the period from January 2, 2000, through October 30, 2016.

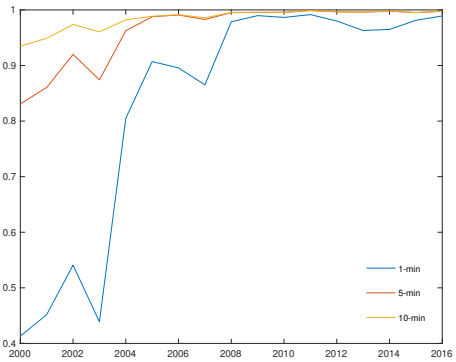
	1-min	5-min	10-min
	Full sample		
5-year note	83.79%	96.48%	98.50%
10-year note	94.12%	98.97%	99.20%
Long term bond	94.78%	99.11 %	99.25%
	Trading days with no activity for more than an hour discarded		
5-year note	84.54%	97.13%	99.11%
10-year note	94.80%	99.59%	99.80%
Long term bond	95.44%	99.72 %	99.84%

Table A.3: Out-of-sample R^2 for two months ahead yield volatility forecasting This table displays R^2_{OoS} measures in percent relative to a RW for all forecasting methods and maturities. In parentheses asymptotic p-values for a one-sided Diebold-Mariano test using the Newey-West variance estimator with automatic lag selection of Andrews (1991). For common factor based forecasters, the label "Initial" indicates that the selection of PCA factors is based on the initial estimation period, and "Recursive" that it is updated every period. The initial estimation period ranges from January 2, 2000, through June 6, 2008, and the out-of-sample period from June 7, 2008, through October 31, 2016.

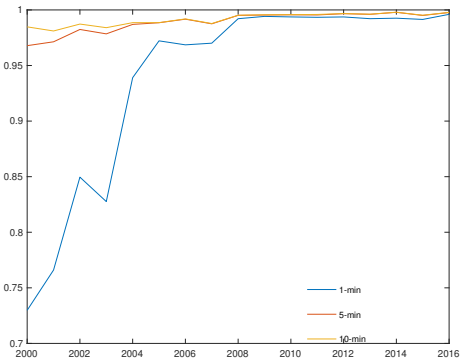
	$\tau = 0.5$	$\tau = 1$	$\tau = 5$	$\tau = 7$
Term structure models				
Vasicek	32.81	42.48	31.66	10.97
	(0.20)	(0.17)	(0.17)	(0.26)
CIR	-30.00	-15.55	4.92	3.08
	(0.71)	(0.62)	(0.44)	(0.44)
$AFNS_0$	-0.77	-128.54	39.98	17.20
	(0.52)	(0.86)	(0.12)	(0.18)
$AFNS_3$	4.29	6.88	38.94	3.78
	(0.44)	(0.40)	(0.13)	(0.45)
$AFNS_3$ (2nd mom)	25.97	-48.89	35.55	16.59
	(0.27)	(0.77)	(0.17)	(0.19)
PCA based forecasters				
PCA_3	7.22	2.37	44.65	11.46
	(0.40)	(0.47)	(0.13)	(0.21)
PCA_4	5.96	0.46	44.17	8.55
	(0.42)	(0.49)	(0.13)	(0.27)
PCA_5	5.44	2.77	44.37	8.48
	(0.43)	(0.47)	(0.12)	(0.28)
PCA_6	-19.31	-12.7	38.12	10.07
	(0.85)	(0.71)	(0.17)	(0.25)
Risk premium based forecasters				
Forward spreads	-0.74	-9.15	46.09	14.15
	(0.51)	(0.60)	(0.10)	(0.20)
Yield spreads	0.08	-9.63	46.52	13.88
	(0.50)	(0.61)	(0.10)	(0.21)
Common factor based forecasters				
Initial	10.64	7.98	27.75	16.79
	(0.35)	(0.40)	(0.28)	(0.13)
Recursive	-5.29	-5.32	34.36	17.25
	(0.62)	(0.60)	(0.20)	(0.16)
Time series models				
HAR	-24.72	-356.3	42.30	-21.15
	(0.97)	(0.87)	(0.14)	(0.67)
Mean-reverting realized GARCH	-43.86	-47.38	29.08	-50.12
	(0.94)	(0.88)	(0.15)	(0.92)

Figure A.1: Liquidity of Treasury futures over time

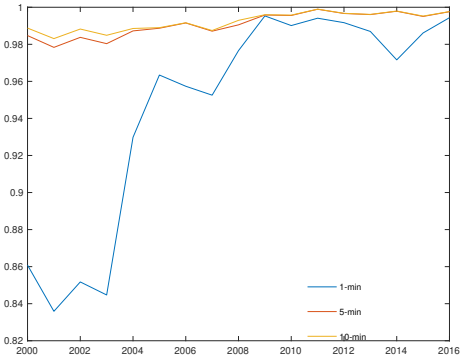
This figure shows the evolution in the daily fraction of high-frequency intervals containing a Treasury futures price observation, by underlying maturity and sampling frequency. The sample spans the period from January 2, 2000, through October 31, 2016.



(a) 5-year note



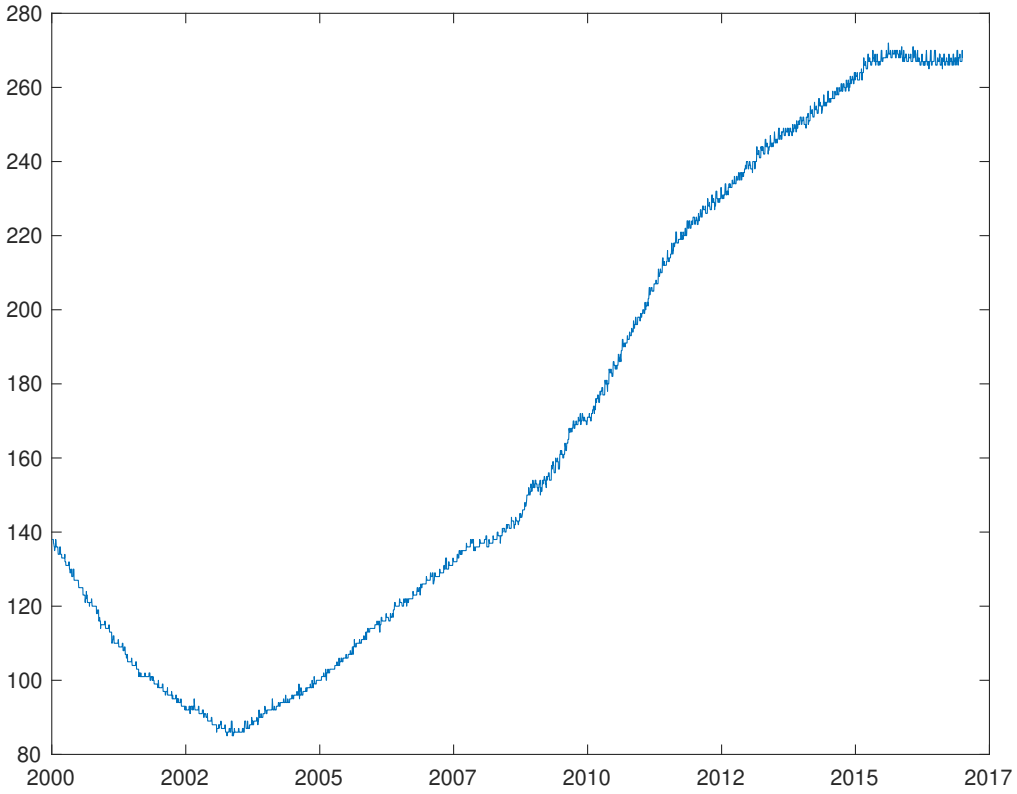
(b) 10-year note



(c) Long term bond

Figure A.2: Number of coupon bonds over time

This figure shows the evolution in the number N_t of coupon bonds with maturity between 3 months and 10 years used in the daily cross-sectional estimations Eq. (34). Callable issues are excluded. The sample spans the period from January 2, 2000, through October 31, 2016.



Research Papers 2021

CREATES
Center for Research in Econometric Analysis of Time Series

2020-15: Carlos Vladimir Rodríguez-Caballero and J. Eduardo Vera-Valdés: Air pollution and mobility in the Mexico City Metropolitan Area, what drives the COVID-19 death toll?

2020-16: J. Eduardo Vera-Valdés: Temperature Anomalies, Long Memory, and Aggregation

2020-17: Jesús-Adrián Álvarez, Malene Kallestrup-Lamb and Søren Kjærgaard: Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans

2020-18: Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman: A statistical model of the global carbon budget

2020-19: Eric Hillebrand, Jakob Mikkelsen, Lars Spreng and Giovanni Urga: Exchange Rates and Macroeconomic Fundamentals: Evidence of Instabilities from Time-Varying Factor Loadings

2021-01: Martin M. Andreasen: The New Keynesian Model and Bond Yields

2021-02: Daniel Borup, David E. Rapach and Erik Christian Montes Schütte: Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

2021-03: Kim Christensen, Mathias Siggaard and Bezirgen Veliyev: A machine learning approach to volatility forecasting

2021-04: Fabrizio Iacone, Morten Ørregaard Nielsen and Robert Taylor: Semiparametric Tests for the Order of Integration in the Possible Presence of Level Breaks

2021-05: Stefano Grassi and Francesco Violante: Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas

2021-06: Gloria González-Rivera, Carlos Vladimir Rodríguez-Caballero and Esther Ruiz Ortega: Expecting the unexpected: economic growth under stress

2021-07: Matei Demetrescu and Robinson Kruse-Becher: Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models

2021-08: Luisa Corrado, Stefano Grassi and Aldo Paolillo: Modelling and Estimating Large Macroeconomic Shocks During the Pandemic

2021-09: Leopoldo Catania, Alessandra Luati and Pierluigi Vallarino: Economic vulnerability is state dependent

2021-10: Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models

2021-11: Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk