Does realized volatility help bond yield density prediction?

Minchul Shin and Molin Zhong*
University of Pennsylvania

This version: November 4, 2013
First version: October, 2012

Abstract

This paper examines the importance of realized volatility in bond yield density prediction. We incorporate realized volatility into a Dynamic Nelson-Siegel (DNS) model with stochastic volatility and evaluate its predictive performance on US bond yield data. When compared to popular specifications in the DNS literature without realized volatility, we find that having this information improves density forecasting performance.

Key words: Dynamic factor model, forecasting, stochastic volatility, term structure of interest rates

JEL codes: C5, G1, E4

*Correspondence: Minchul Shin: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. E-mail: mincshin@sas.upenn.edu. Molin Zhong: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. E-mail: molin@sas.upenn.edu. We are grateful for the advice of Frank Diebold and Frank Schorfheide. We also thank Luigi Bocola, Xu Cheng, Frank DiTraglia, Nikolaus Hautsch, Kyu Ho Kang, Andrew Patton, Neil Shephard, Dongho Song, and Allan Timmermann and participants in the Penn econometrics club, International Symposium on Forecasting 2013, and OMI-SoFiE Financial Econometrics Summer School 2013 for their comments.

1 Introduction

Existing term structure models seldom beat the random walk in terms of point prediction. This fact, however, may not carry over to bond yield density prediction, as first pointed out by Egorov, Hong, and Li (2006). Lack of conditional mean predictability does not imply the same for the conditional variance. Researchers have found that time-varying volatility exists and mean reverts in US government bond yields. This has led Egorov, Hong, and Li (2006) to consider affine term structure models, Koopman, Mallee, and Van der Wel (2010) to introduce a factor GARCH in the Diebold and Li (2006) Dynamic Nelson Siegel (DNS) model, Hautsch and Ou (2012) and Hautsch and Yang (2012) to incorporate stochastic volatility in the DNS model, and Carriero, Clark, and Marcellino (2013) to look at a shrinkage-based vector autoregression with stochastic volatility, all in the hopes of better forecasting bond yield densities.

In this paper, we take a different approach: introducing volatility proxy data. Specifically, we have a DNS model for bond yields with stochastic volatility at the monthly frequency and incorporate monthly bond yield realized volatility. We find that the higher frequency movements of the yields in the realized volatility data contain valuable information for the stochastic volatility and lead to significantly better density predictions, especially in the short term.

The DNS class of models uses latent level, slope, and curvature factors to drive the intertemporal movements of the yield curve. This reduces the high-dimensional yields to be driven by just three factors. The level of the yield curve has traditionally been linked to inflation expectations while the slope to the real economy. Our preferred specification introduces stochastic volatility on these latent factors. This leads to a nice interpretation of the stochastic volatility as capturing the uncertainty surrounding well-understood aspects of the yield curve. It also reduces the dimension of modeling the time-varying volatility of the yield curve. We add realized volatility by linking this data to the model-implied conditional

volatility through a measurement equation. This brings the factor structure in modeling the yields to the second moments as well.

We then compare this specification to several others in the DNS framework, including random walk dynamics for the factors and stochastic volatilities and stochastic volatility on the yield measurement equation. In a forecasting horserace on US bond yields, our preferred specification features slight improvements in the point forecast performance and significant gains in the density forecast performance. We also find that allowing time-varying volatility is important for density prediction, especially in the short run. Unlike conditional mean dynamics, modeling volatility as stationary processes rather than random walks leads to better predictive performance. Furthermore, having stochastic volatility on the factor equation better captures the time-varying volatility in the bond yield data when compared to stochastic volatility on the measurement equation. Finally, in addition to the standard forecast evaluation criterion, we also evaluate our models along economically meaningful dimensions in the forms of forecasting empirical level, slope, and curvature.

Our paper relates to the literature in three main areas.

First, we contribute to a large literature on bond yield forecasting. Most of the work has been done on point prediction (see Diebold and Rudebusch (2012) and Duffee (2012) for excellent surveys). There has been, however, a growing interest in density forecasting. Egorov, Hong, and Li (2006) were the first to evaluate the joint density prediction performance of yield curve models. They overturn the point forecasting result of the superiority in random walk forecasts and find that affine term structure models perform better when forecasting the entire density, especially the conditional variance and kurtosis. Hautsch and Ou (2012) and Hautsch and Yang (2012) add stochastic volatility to the DNS model by considering an independent AR(1) specification for the log volatilities of the latent factors. They do not do formal density prediction evaluation of the model, but give suggestive results of the possible improvements in allowing for time-varying volatility. Carriero, Clark, and Marcellino (2013) find that using priors from a Gaussian no-arbitrage model in the context of a VAR with

stochastic volatility improves short-run density forecasting performance. Building on this previous work, we introduce potentially highly accurate volatility information into the model in the form of realized volatility. We also extend the forecast evaluation by considering joint density forecasting in the forms of empirical level, slope, and curvature.

Second, we also add to a growing literature on including realized volatility information in bond yield models. Andersen and Benzoni (2010) and Christensen, Lopez, and Rudebusch (2011) view realized volatility as a benchmark on which to compare the fits of affine term structure models. Cieslak and Povala (2013) are interested in using realized covariance to better extract stochastic volatility and linking the stochastic volatility to macroeconomic and liquidity factors. These papers focus on in-sample investigations of incorporating realized volatility in bond yield models. Our paper, on the other hand, considers the improvement from using realized volatility in out-of-sample point and density prediction.

Finally, our paper relates to work started by Barndorff-Nielsen and Shephard (2002) in incorporating realized volatility in models with time-varying volatility. Takahashi, Omori, and Watanabe (2009) use daily stock return data in combination with high-frequency realized volatility to more accurately estimate the stochastic volatility. Maheu and McCurdy (2011) show that adding realized volatility directly into a model of stock returns can improve density forecasts over a model that only uses level data, such as the EGARCH. Jin and Maheu (2013) propose a model of stock returns and realized covariance based on time-varying Wishart distributions and find that their model provides superior density forecasts for returns. There also exists work adding realized volatility in observation-driven volatility models (Shephard and Sheppard (2010) and Hansen, Huang, and Shek (2012)). As opposed to the other papers, we consider a dynamic factor model with stochastic volatility on the factor equation and use the realized volatility to help in the extraction of this stochastic volatility. In this sense, we bring the factor structure in the conditional mean to the conditional volatility as well[1]. Furthermore, we are the first paper to investigate the implications of realized volatility on

[1]Cieslak and Povala (2013) have a similar framework in a no-arbitrage term structure model

bond yield density predictability.

In section 2, we introduce our preferred specification and competitor specifications. We discuss the data in section 3. We present our estimation and forecast evaluation methodology in section 4. In section 5, we present in-sample and out-of-sample results. We conclude in section 6.

2 Model

We introduce the Dynamic Nelson Siegel model with stochastic volatility (DNS-SV) proposed by Hautsch and Ou (2012), Hautsch and Yang (2012), and Bianchi, Mumtaz, and Surico (2009). Then, we discuss the incorporation of realized volatility information into this framework. Finally, we consider alternatives to our main approach.

2.1 The Dynamic Nelson-Siegel model and time-varying bond yield volatility

Denote $y_t(\tau)$ as the continuously compounded yield to maturity on a zero coupon bond with maturity of τ periods at time t. Following Diebold and Li (2006), we consider the factor model for the yield curve,

$$y_t(\tau) = f_{l,t} + f_{s,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + f_{c,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \epsilon_t(\tau), \quad \epsilon_t \sim N(0,\ Q) \qquad (1)$$

where $f_{l,t}, f_{s,t}$ and $f_{c,t}$ serve as latent factors and ϵ_t is a vector that collects idiosyncratic component $\epsilon_t(\tau)$ for all maturities. As is well documented in the literature, the first factor mimics the level of the yield curve, the second the slope, and the third the curvature. While some authors have estimated the λ, we fix it at 0.0609, noting from Diebold and Li (2006) and others that the value does not move around too much across time and that its estimation

does not seem to affect the results. We assume that the Q matrix is diagonal. This leads to the natural interpretation of a few common factors driving the comovements in a large number of yields. All of the other movements in the yields are considered idiosyncratic. As suggested by Diebold and Li (2006), we model the dynamic factors as independent univariate first-order autoregressive processes, given by,

$$f_{i,t} = \mu_{f,i}(1 - \phi_{f,i}) + \phi_{f,i} f_{i,t} + \eta_{i,t}, \quad \eta_t \sim N(0,\ H_t) \tag{2}$$

for $i = l, s, c$. η_t is a vector that collects the innovations to each factor and its variance-covariance matrix H_t potentially varies over time. We also assume that idiosyncratic shocks ϵ_t and factor shocks η_t are independent. The independent factor specification of the setup means that the movements of the factors are unrelated to each other. While this may seem as a tight restriction at first blush, Diebold and Rudebusch (2012) point out that the assumption does not seem poor in so far as the factors are related to the principal components of the yield curve. Following Hautsch and Ou (2012), (Hautsch and Yang, 2012) and Bianchi, Mumtaz, and Surico (2009), we model the logarithm of the variance of the shocks to the factor equation as AR(1) processes,

$$h_{i,t} = \mu_{h,i}(1 - \phi_{h,i}) + \phi_{h,i} h_{i,t-1} + \nu_{i,t}, \quad \nu_{i,t} \sim N(0,\ \sigma^2_{h,i}) \tag{3}$$

for $i = l, s, c$. $\exp(h_{i,t})$ corresponds to the ith diagonal element of the variance-covariance matrix H_t. Since the conditional mean dynamics of the factors are specified as independent from each other, it makes sense to model the stochastic volatilities of the factor innovations as independent. We call this specification the DNS-SV model (Dynamic Nelson-Siegel with stochastic volatility).

2.2 DNS-RV

We claim that by using high-frequency data to construct realized volatilities of the yields it is possible to aid in the extraction of the stochastic volatilities governing the level, slope, and curvature of the DNS-SV model. Using realized volatility to augment our algorithm should make estimation of the stochastic volatility parameters more accurate and produce a superior predictive distribution. Crucially, we need to find an appropriate linkage between our volatility proxy - realized volatility - and the stochastic volatility in the model. Given the definition of the model-implied conditional volatility, we propose

$$\begin{aligned} RV_t \approx Var_{t-1}(y_t) &= diag(\Lambda_f H_t \Lambda_f' + Q) \\ &= diag(\tilde{\Lambda}_f \tilde{H}_t \tilde{\Lambda}_f' + Q) \end{aligned} \tag{4}$$

where RV_t is the realized volatility of bond yields which has the same dimension as the bond yield vector y_t and Λ_f is factor loading matrix given by equation 1. We write the logarithm of volatility in deviation form $\tilde{h}_{i,t} = h_{i,t} - \mu_{h,i}$ for $i = l, s, c$. Then $\tilde{H}_t$ is a 3×3 diagonal matrix with each element corresponding to $e^{\tilde{h}_{i,t}}$ and $\tilde{\Lambda}_f = \Lambda_f[e^{\mu_l/2}, e^{\mu_s/2}, e^{\mu_c/2}]'$. The second equality in equation 4 comes from the change in notation. [2] Insofar as realized volatility provides an accurate approximation to the true underlying conditional time-varying volatility, equation 4 is the one that links this information to the model.

With measurement error, one can view equation 4 as a nonlinear measurement equation. In principle, we have several tools to handle this nonlinearity, including the particle filter. To keep estimation computationally feasible, we choose to take a first order Taylor approximation of the logarithm of this equation around $\tilde{h}_t = 0$ with respect to h_t. This leads to a set of linear measurement equations that links the realized volatility of the bond yields and

[2] This strategy of linking an observed volatility measure to the model is also used in other papers (Maheu and McCurdy (2011) in a univariate model and Cieslak and Povala (2013) in a multivariate context).

the underlying factor volatility,

$$\log(RV_t) = \beta + \Lambda_h \tilde{h}_t + \zeta_t, \quad \zeta_t \sim N(0, S). \tag{5}$$

In the estimation, we add in measurement error ζ_t. We view equation 5 as an approximation to equation 4. We call this new model the Dynamic Nelson Siegel with realized volatility (DNS-RV) model. The only difference between this model and DNS-SV comes from augmenting equation 1 with a new measurement equation 5. This equation has a constant β and a factor loading Λ_h. The parameter β comes from the linearization while we can interpret Λ_h as a loading for the factor volatility used to reduce the dimension of the volatility data, $\log(RV_t)$. This very naturally extends the dynamic factor model, which transforms high-dimensional data (y_t) into a few number of factors (f_t) via the factor loading matrix Λ_f. The volatility factor loading (Λ_h) is a function of other model parameters (Λ_f, μ_h) with the functional form given by the linearized version of equation 4.[3] We can view this as a model-consistent restriction on the linkage between the conditional volatility of observed data, approximated by $\log(RV_t)$, and the factor volatility h_t. For the same reasons as in the baseline DNS-SV model, we set the S matrix to be diagonal. These are interpreted as idiosyncratic errors, and we therefore do not model them to be contemporaneously or serially correlated.

In summary, the DNS-RV model introduces a new measurement equation into the state space of the DNS-SV model.

(Measurement equation)

$$\begin{aligned} y_t &= \Lambda_f f_t + \epsilon_t, \quad \epsilon_t \sim N(0,\ Q) \\ \log(RV_t) &= \beta + \Lambda_h \tilde{h}_t + \zeta_t, \quad \zeta_t \sim N(0, S), \end{aligned} \tag{6}$$

[3]Detailed formulas for the volatility factor loading matrix can be found in the appendix.

Table 1 Model Specifications

Label	Factors (level, slope, curvature)	Conditional variance	Realized volatility
RW-C	Random walk	Constant	Not used
RW-SV	Random walk	log AR(1) in each factor	Not used
RW-RV	Random walk	log AR(1) in each factor	Used
DNS-C	Independent AR(1)	Constant	Not used
DNS-SV	Independent AR(1)	log AR(1) in each factor	Not used
DNS-RV	Independent AR(1)	log AR(1) in each factor	Used
RW-SV-RW	Random walk	Random walk	Not used
DNS-SV-RW	Independent AR(1)	Random walk	Not used
DNS-ME-SV	Independent AR(1)	log AR(1) in measurement equation	Not used
DNS-ME-RV	Independent AR(1)	log AR(1) in measurement equation	Used

Note: We list the specifications for the DNS model considered in this paper.

(Transition equation)

$$\begin{aligned} f_{i,t} &= \mu_{i,f}(1-\phi_{i,f}) + \phi_{i,f} f_{i,t} + \eta_t, \quad \eta_t \sim N(0,\ H_t) \\ h_{i,t} &= \mu_{i,h}(1-\phi_{i,h}) + \phi_{i,h} h_{i,t-1} + \nu_{i,t}, \quad \nu_{i,t} \sim N(0,\ \sigma^2_{i,h}) \end{aligned} \tag{7}$$

for $i = l, s, c$. In our application, both observed bond yields (y_t) and realized volatilities ($\log(RV)_t$) are 17×1 vectors. Moreover, both sets of variables have a factor structure with dynamics following the transition equations.

2.3 Alternative specifications

We have four classes of alternative specifications to compare forecasts to our baseline model. We briefly introduce them in this section and list all specifications considered in the paper in Table 1.

2.3.1 Dynamic Nelson Siegel (DNS-C)

The first model is the standard Diebold-Li DNS model discussed at the beginning of the paper. It does not allow for stochastic volatility. This model has been shown to forecast the level of bond yields quite well, at times beating the random walk model of yields.

2.3.2 Dynamic Nelson Siegel-Stochastic Volatility (DNS-SV)

The second is the DNS-SV model that adds stochastic volatility to the transition equation. It is summarized at the beginning of this section. By allowing for stochastic volatility, this model should improve upon the standard DNS model especially in the second moments, as it can capture the time-varying volatility present in the bond yield data.

2.3.3 Dynamic Nelson Siegel-Random Walk (RW)

Bond yield forecasting research (e.g. van Dijk, Koopman, van der Wel, and Wright (2013) and references therein) has shown that random walk specifications of the yield curve generally perform quite well. Oftentimes, the no-change forecast from a current period does best among a large group of forecasting models. It is in this sense that bond yield forecasting is difficult. Given these results, we also augment the DNS-C, DNS-SV, and DNS-RV model classes with random walk parameterizations of the factor processes.

Macroeconomic research (Cogley and Sargent (2005) and Justiniano and Primiceri (2008)) often specifies stochastic volatility as following a random walk. Doing so reduces the number of parameters estimated while also providing a simple no-change forecast benchmark for time-varying volatility. As long-horizon bond yield volatility links with macroeconomic volatility, we also have random walk specifications for the stochastic volatilities.

2.3.4 Dynamic Nelson Siegel-Measurement Error Stochastic Volatility (+ Realized Volatility) (DNS-ME)

Koopman, Mallee, and Van der Wel (2010) argue that putting the time-varying conditional volatility on the measurement errors provides an improvement for the in-sample fit of the DNS class of models. To evaluate whether these results extend to forecasting as well, we model independent AR(1) specifications for the measurement error stochastic volatilities. While following this strategy greatly increases the number of parameters estimated, it could

improve forecasting as each yield has its own stochastic volatility process. Therefore, as opposed to a time-varying H_t and constant Q matrix in the DNS-SV and DNS-RV setups, now H is constant while Q_t has stochastic volatility.

$$Q_t = \mathbf{Diagmat}(e^{\mathbf{h_t}}) \tag{8}$$

$$h_{i,t} - \mu_{h,i} = \phi_{h,i}(h_{i,t-1} - \mu_{h,i}) + \nu_{i,t}, \quad \nu_{i,t} \sim N(0,\ r_i), \tag{9}$$

for $i = 1, ..., N$ where N is the number of bond yields in the observation equation. $\mathbf{h_t}$ is a vector that collects all stochastic volatilities in measurement errors and $\mathbf{Diagmat}(\cdot)$ turns a vector into a diagonal matrix. Q_t remains a diagonal matrix, as equation 8 shows. We again model the logarithm of the variances as independent first order autoregressive processes.

We also consider incorporating realized volatility information into this model. Doing so leads to the following relationship

$$RV_t \approx Var_{t-1}(y_t) = diag(\Lambda_f H \Lambda_f' + Q_t). \tag{10}$$

As before, we do a first order Taylor approximation of the logarithm of this equation. We also add in a measurement error for estimation. However, in contrast to the DNS-RV model, we link each element in $\log(RV_t)$ with individual stochastic volatility $h_{i,t}$.

3 Data

We use a panel of unsmoothed Fama and Bliss (1987) U.S. government bond yields at the monthly frequency with maturities of 3, 6, 9, 12, 15, 18, 21 months and 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10 years from January 1981 to December 2009. This dataset is provided by Jungbacker, Koopman, and van der Wel (2013)[4]. To construct monthly realized volatility series, we use

[4] <http://qed.econ.queensu.ca/jae/datasets/jungbacker001/>

daily U.S. government bond yield data with the same maturities from January 2, 1981 to December 30, 2009 taken from the Federal Reserve Board of Governors with the methodology of Gürkaynak, Sack, and Wright (2007) [5] [6].

We construct the realized variance of each month's yields using daily bond yield data. The formula for realized variance at time t is

$$RV_t = \sum_{d=1}^{D} \left(\Delta y_{t+\frac{d}{D}} \right)^2 .$$

where D is the number of daily data in one time period t. This formula converges in probability to the true conditional variance as the sampling frequency goes to infinity under assumptions laid out in Andersen, Bollerslev, Diebold, and Labys (2003). Usually, there are around 21 days in each month, with less depending upon the number of holidays in a month that fall on normal trading days.

We use daily data to construct our realized volatilities for a few reasons. First, we want to use realized volatility information starting in 1981 to use a sample period similar to other bond yield forecasting studies. The availability of higher-frequency intraday data begins much later. For instance, Cieslak and Povala (2013) start their estimation in 1992 for specifically this reason. Second, the month-to-month volatility movements we want the volatility proxy to capture do not necessitate using ultra-high frequency data. Finally, while results may improve with higher frequency data, we show that positive effects are present even with lower frequency realized volatility.

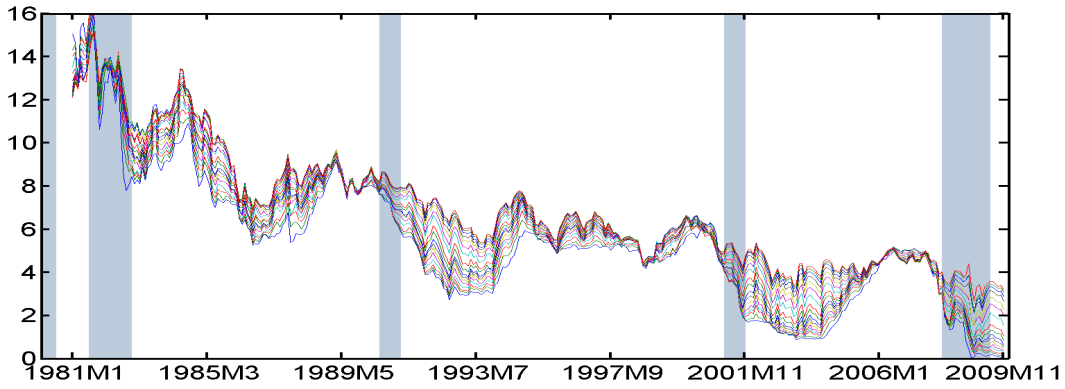
Figure 1 plots the time series of monthly U.S. government bond yields and logarithm of realized volatilities. All yields exhibit a general downward trend from the start of the sample. For around the first 25 months, the realized volatility seems quite high and exhibits large time

[5] <http://www.federalreserve.gov/econresdata/researchdata/feds200628_1.html>

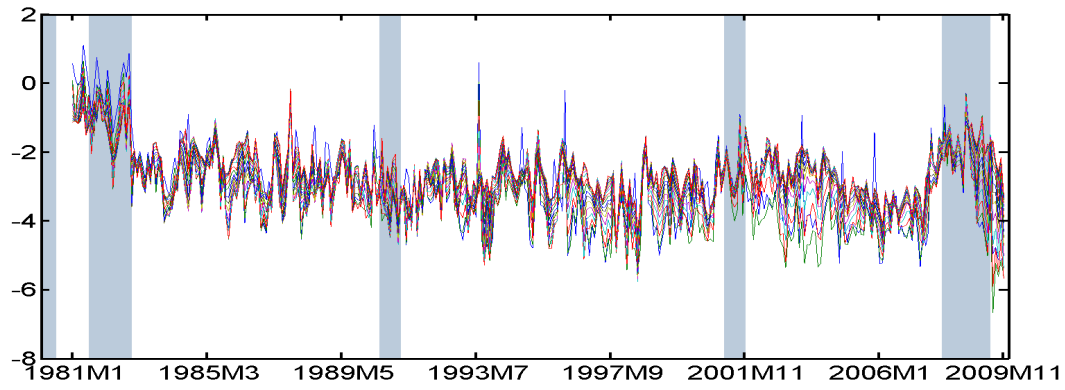
[6] Because most papers in the literature estimate the DNS model with unsmoothed Fama and Bliss data, we generate and evaluate predictions for the unsmoothed Fama and Bliss data. Unfortunately, this data is only available at the monthly frequency. Even though the two datasets use different methodologies, monthly yield data based on the daily bond yield from Gürkaynak, Sack, and Wright (2007) is very close to the one based on the unsmoothed Fama and Bliss method.

Figure 1 U.S. Treasury Yields

(a) Yields (monthly, annualized %)



(b) Yield Realized Volatilities (Monthly, Log)



Notes: We present monthly U.S. Treasury yields with maturities of 3, 6, 9, 12, 15, 18, 21 months and 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10 years over the period January 1981 – November 2009. Monthly yields are constructed using the unsmoothed Fama-Bliss method. Monthly yield realized volatilities are constructed based on daily yields using Wright's dataset. Blue shaded bars represent NBER recession dates.

variation. After around 1983, yield volatility dies down and largely exhibits only temporary spikes in volatility. For a period of 2 years starting in 2008, the realized volatility picks up across all yields. We attribute this to the financial crises. Another interesting feature of log realized volatility is that it shows large autocorrelation. Its first-order autocorrelation coefficients range from 0.59 to 0.69 and the 12th-order autocorrelation coefficients range from 0.20 to 0.31.[7] This means that realized volatility data could help even for the long-horizon

[7]We provide tables in the supporting material for the descriptive statistics of monthly realized volatility of bond yields.

Table 2 Variance explained by the first five principal components (%)

	Yield	log(RV)
pc 1	98.16	84.30
pc 2	99.84	94.62
pc 3	99.95	98.53
pc 4	99.98	99.46
pc 5	99.98	99.85

Notes: Numbers in the table are the percentage of total variance explained by the first five principal components for US Yield data and log(RV). log(RV) is the logarithm of the monthly realized volatility constructed based on the daily US yield data.

forecasts that we consider in this paper.

Both the yields and realized volatilities do seem to exhibit a factor structure, meaning that each set of series co-move over time. In fact, a principal components analysis shows that the first three factors for yields explain 99.95% of the variation in the US yield curve. The first three factors for realized volatilities explain 98.53% of the variation (Table 2). Even though the fact that U.S. bond yield can be explained by the first few principal components is well documented, it is interesting that the same feature carries over for the realized volatility of U.S. bond yields.

4 Estimation/Evaluation Methodology

4.1 Estimation

We perform a Gibbs sampling Markov Chain Monte Carlo algorithm for 15000 draws. We keep every 5th draw and burn in for the first 5000 draws. Due to our linearization approximation in introducing realized volatility, all specifications that we consider can be sampled by using the method for the stochastic volatility state space model developed in Kim, Shephard, and Chib (1998). Details of the state space representation and the estimation procedure can be found in the appendix. Details on the prior can be found in the appendix as well, although we comment that our choice of prior is loose and does not impact the estimation results.

We highlight the difference in estimation procedure due to the additional measurement equation for the realized volatility. Roughly speaking, there are two sources of information for the latent volatility factor h_t. The first source is from the latent factor, f_t. To see this, one can transform the transition equation for f_t as the following.

$$\log\left((f_{i,t} - \mu_{i,f}(1-\phi_{i,f}) - \phi_{i,f} f_{i,t})^2\right) = h_{i,t} + \log(x_{i,t}^2), \quad x_{i,t} \sim N(0,1)$$

for $i = l, s, c$. This is common to both DNS-SV and DNS-RV. The second source is from equation 6 which relates $\log(RV_t)$ with h_t and is unique to DNS-RV. Following Kim, Shephard, and Chib (1998), we approximate $\log(x_{i,t}^2)$ with mixture of normals. Then, conditional on other parameters and data, extraction of h_t amounts to running a simulation smoother in conjunction with the Kalman filter with and without equation 6.

4.2 Forecast evaluation

We consider model performance along both the point and density forecasting dimensions. The appendix contains further details on the Bayesian simulation algorithm we use to generate the forecasts. We begin forecasting on February 1994 and reestimate the model in an expanding window and forecast moving forward two months at a time. For every forecast run at a given time t, we forecast for all yields in our dataset and for horizons ranging from 1 month to 12 months ahead. This leads to a total of 94 repetitions.

Point forecast

To evaluate the point prediction, we use the Root Mean Square Forecast Error (RMSE) statistic.

$$RMSE_{\tau,ho}^{M} = \sqrt{\frac{1}{F}\sum\left(\hat{y}_{t+ho}^{M}(\tau) - y(\tau)_{t+ho}\right)^2} \tag{11}$$

Call the yield τ forecast at horizon ho made by model M as $\hat{y}^M_{t+ho}(\tau)$ and $y(\tau)_{t+ho}$ the realized value of the yield at time $t+ho$. F is the number of forecasts made. Then, equation 11 provides the formula for the RMSE. To gauge whether there are significant differences in the RMSE, we use the Diebold and Mariano (1995) t-test of equal predictive accuracy.

Density forecast

The log predictive score (Geweke and Amisano (2010)) gives an indication of how well a model performs in density forecasting.

$$LPS^M_{\tau,ho} = \frac{1}{F}\sum \log p(y_{t+ho}(\tau)|y^t, M) \tag{12}$$

$p(y_{t+ho}(\tau)|y^t, M)$ denotes the ho-step ahead predictive distribution of yield τ generated by model M given time t information. Following Carriero, Clark, and Marcellino (2013), we estimate the log predictive density by a kernel density estimator using MCMC draws for parameters and latent states and compute the p-value for the Amisano and Giacomini (2007) t-test of equal means to gauge whether there exist significant differences in the log predictive score.

5 Results

We first present in-sample results of the model, focusing on time-varying volatility. Then, we move to point and density forecasting results. Finally, we discuss our empirical factor prediction exercise.

5.1 In-sample

We first present the full sample estimation from January 1981 - November 2009. We focus on how adding realized volatility information alters the model. Adding in second-moment

Table 3 Posterior Estimates of Parameters on h_t Equation

	DNS-SV			DNS-RV		
	5%	50%	95%	5%	50%	95%
$\mu_{h,l}$	-4.40	-2.56	-1.48	-4.19	-3.92	-2.21
$\mu_{h,s}$	-3.41	-2.30	-1.67	-3.22	-2.85	-1.94
$\mu_{h,c}$	-1.48	-0.96	-0.47	-2.18	-1.88	-1.39
$\phi_{h,l}$	0.93	0.98	0.999	0.58	0.66	0.73
$\phi_{h,s}$	0.92	0.97	0.995	0.57	0.64	0.72
$\phi_{h,c}$	0.81	0.92	0.98	0.39	0.49	0.60
$\sigma_{h,l}$	0.01	0.03	0.08	0.41	0.75	0.92
$\sigma_{h,s}$	0.01	0.04	0.10	1.27	1.72	2.64
$\sigma_{h,c}$	0.02	0.09	0.27	1.30	1.78	4.30

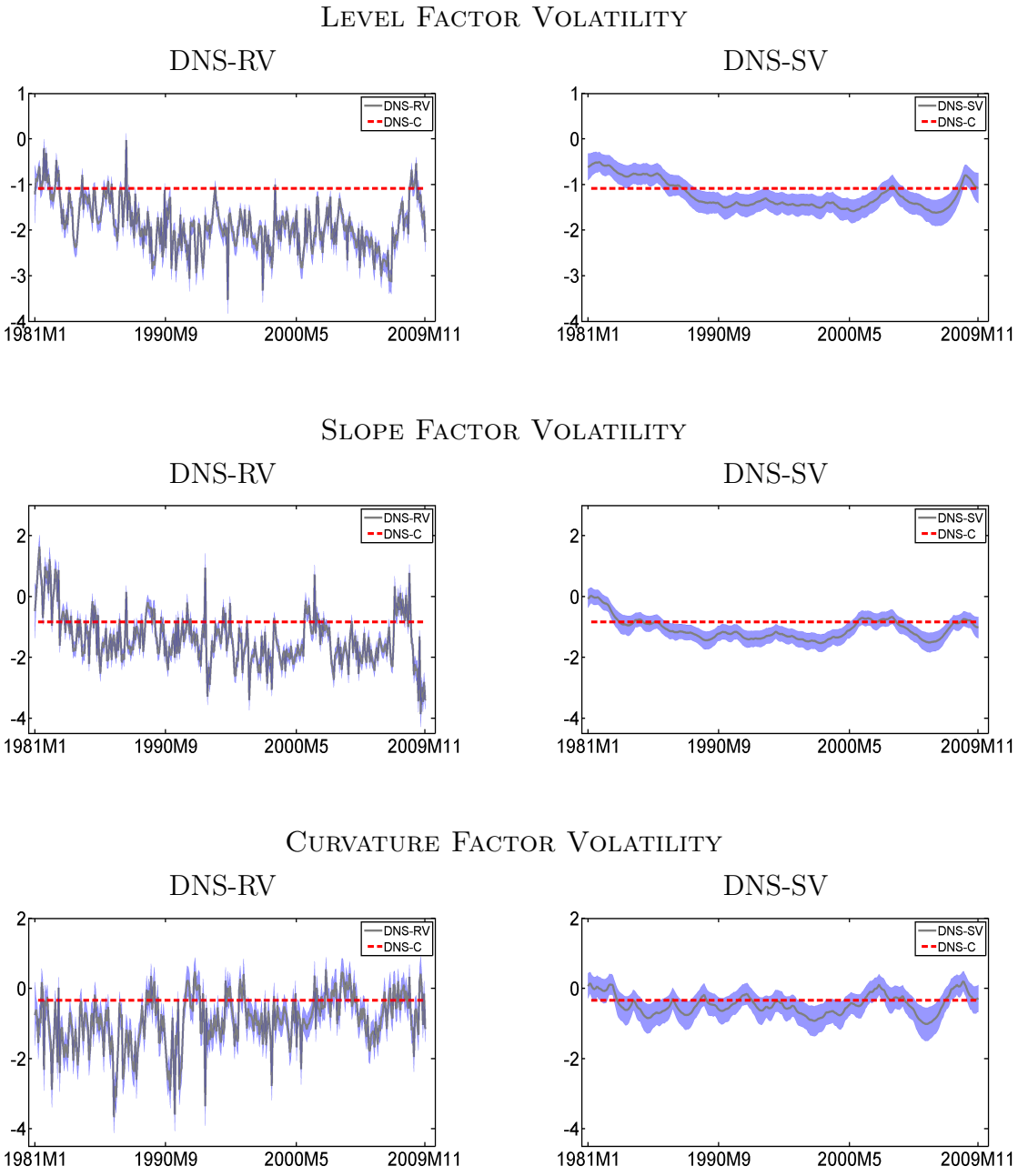
Notes: Posterior moments are based on estimation sample from January 1981 to November 2009.

information does not significantly change conditional mean dynamics, so we relegate our discussion of the extracted factors to the supporting material[8]. Our extracted factors are similar to those found in Diebold and Li (2006).

The stochastic volatility dynamics deserve some more precise discussion. Figure A-1 shows the volatility estimate from the DNS-C, DNS-SV, and DNS-RV models. The fluctuations of the extracted stochastic volatilities in both the DNS-SV and DNS-RV models show that there exists conditional time-varying volatility in the data. Relative to the DNS-SV specification, adding in realized volatility data makes the extracted stochastic volatility much less persistent and more variable. This leads to a lower autoregressive parameter and higher innovation standard deviation estimates for all of the stochastic volatility processes (Table 3). The DNS-RV model also delivers lower stochastic volatility mean estimates. These differences lead to differences in forecasting. For example, the lower autoregressive parameter in the DNS-RV model means that it predicts faster mean reversion of the stochastic volatilities relative to the DNS-SV model and the lower long-run mean estimate implies that the DNS-RV model produces a tighter density prediction in the long run. The smoothed stochastic volatilities from the DNS-SV model, however, generally captures the low-frequency movements of those

[8]All other in-sample estimation results are in the supporting material

Figure 2 Stochastic volatility for bond yield factors



Notes: Posterior median of log stochastic volatility (h_t) for bond yield factors from DNS-RV (left column) and DNS-SV (right column). Red dotted line is volatility level estimated from DNS-C. Blue band is 80% credible interval. Estimation sample is from January 1981 to November 2009.

from the DNS-RV model.

We argue that this difference in the stochastic volatilities matters for density forecasting. The high-frequency data used to construct the realized volatilities brings information that the low-frequency monthly yield data misses. By having more accurate estimates of the current level of time-varying volatility and process parameters, the DNS-RV model both starts off forecasting at a more accurate point and better captures the dynamics of the data moving forward.

5.2 Point prediction

Table 4 shows the RMSE of selected maturities for 1, 3, 6, and 12-step ahead predictions. The second column has the calculated RMSE values for the RW-C model. All other values reported are ratios relative to the RW-C RMSE. Values below 1 indicate superior performance relative to the random walk benchmark. Stars in the table indicate significant gains relative to the RW-C model. As expected, RMSE increases as the forecasting horizon lengthens. The models with random walk dynamics in the factors do well for short-horizon forecasts but deteriorate when compared to the stationary models as the prediction horizon lengthens. In general, all RMSE values have numbers close to 1, reproducing the well-known result in the bond yield forecasting literature on the difficulty in beating the no-change forecast. As alluded to in the previous section, adding in time-varying second moments does not largely impact point predictions, although the DNS-RV model forecasts middle maturities well across all horizons. For 12-month horizon forecasts, the DNS-C and DNS-ME-RV models also do well for short maturity yields.

5.3 Density prediction

Table 5 shows the density evaluation result in terms of log predictive score. Similar to the RMSE table, the RW-C column gives the value of the log predictive score for the random

Table 4 RMSE comparison

Maturity	RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
1-step-ahead prediction										
3	0.267	0.992	1.011	1.004	1.034**	1.007	**0.991**	1.038**	1.071	0.953**
12	**0.229**	1.005	1.002	1.076**	1.055**	1.003	1.002	1.060**	1.070**	1.072**
36	0.274	1.001	0.996	1.015	1.020	**0.990**	1.001	1.025	1.010	1.012
60	0.274	1.000	0.995	1.003	1.010	**0.987**	1.000	1.013	1.002	1.004
120	0.277	1.001	0.996	0.988	0.989	0.999	1.000	0.988	**0.986**	0.997
3-step-ahead prediction										
3	**0.506**	1.002	1.011	1.036	1.066**	1.018	1.002	1.069**	1.055	1.024
12	0.537	1.002	**0.999**	1.072	1.073**	**0.999**	1.002	1.077**	1.064	1.063
36	0.580	1.000	0.998	1.022	1.036	**0.990**	0.999	1.040	1.018	1.018
60	0.551	1.000	0.998	1.004	1.017	**0.986**	0.999	1.021	1.005	1.004
120	0.489	1.000	0.998	**0.982**	0.987	0.995	1.000	0.988	0.987	0.987
6-step-ahead prediction										
3	0.932	1.000	1.002	1.003	1.047	1.007	**0.999**	1.048	1.010	1.000
12	0.915	1.001	1.001	1.040	1.065	**0.999**	1.000	1.068	1.036	1.034
36	0.881	1.001	1.001	1.017	1.041	**0.987**	1.000	1.046	1.011	1.011
60	0.819	1.000	0.998	0.998	1.016	**0.978**	1.000	1.020	0.997	0.997
120	0.665	0.999	0.992*	**0.977**	0.982	0.980	0.998	0.986	0.981	0.980
12-step-ahead prediction										
3	1.592	1.002*	0.999	**0.936****	1.013	0.991	1.000	1.015	0.943**	0.940*
12	1.476	1.001	0.999	**0.975**	1.034	0.991	1.000	1.036	0.976	**0.975**
36	1.242	1.001	1.001	0.997	1.043	**0.979****	1.001	1.046	0.993	0.993
60	1.069	1.002	1.000	0.997	1.031	**0.967***	1.001	1.033	0.994	0.994
120	0.833	1.002	0.998	0.976	0.994	**0.966**	1.001	0.994	0.980	0.976

Notes: The first column shows the RMSE based on the RW-C. Other columns show the relative RMSE compared to the first column. The RMSE from the best model for each variable and forecast horizon is in bold letter. Units are in percentage points. Divergences in accuracy that are statistically different from zero are given by * (10%), ** (5%), *** (1%). We construct the p-values based on the Diebold and Mariano (1995) t-statistics with a variance estimator robust to serial correlation using a rectangular kernel of $h-1$ lags and the small-sample correction proposed by Harvey, Leybourne, and Newbold (1997).

walk case while the numbers for the other models are differences relative to that column. A higher value indicates larger log predictive score and better density forecasting results. We present p-values based on Amisano and Giacomini (2007) comparing the hypothesis of equal log predictive score of the DNS-RV with those of the alternative models in table 6. As opposed to the point prediction results, three interesting findings emerge when we consider the log predictive score.

First, for the short-run horizon, having realized volatility gives significant gains in density prediction. This is on top of a large improvement in log predictive score from adding stochastic volatility, which Carriero, Clark, and Marcellino (2013) find. Table 6 shows that the DNS-RV model has significantly higher log predictive score values for one- and three-

Table 5 Log predictive score comparison

Maturity	RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
1-step-ahead prediction										
3	-0.538	0.205	0.343	0.008	0.198	**0.375**	0.181	0.173	0.048	0.102
12	-0.357	0.211	0.304	-0.009	0.195	**0.316**	0.189	0.172	0.012	0.013
36	-0.318	0.129	0.158	0.000	0.118	**0.176**	0.111	0.104	0.022	0.017
60	-0.265	**0.109**	0.077	-0.001	0.098	0.105	0.098	0.088	0.019	0.015
120	-0.248	**0.110**	0.111	0.011	**0.121**	0.110	0.106	0.108	0.048	0.037
3-step-ahead prediction										
3	-1.061	0.183	0.302	0.016	0.165	**0.322**	0.151	0.133	0.036	0.061
12	-0.985	0.107	0.195	-0.012	0.064	**0.196**	0.089	0.045	0.007	0.010
36	-0.951	0.027	0.053	-0.002	-0.002	**0.067**	0.013	-0.018	0.009	0.007
60	-0.882	0.013	0.016	0.003	-0.011	**0.035**	0.007	-0.015	0.011	0.010
120	-0.779	0.033	0.020	0.014	0.033	0.031	**0.036**	**0.038**	0.027	0.029
6-step-ahead prediction										
3	-1.485	0.048	0.157	0.028	0.023	**0.159**	0.033	0.005	0.038	0.053
12	-1.417	0.010	0.065	0.000	-0.058	**0.066**	0.000	-0.082	0.008	0.012
36	-1.342	-0.017	0.021	0.001	-0.077	**0.030**	-0.039	-0.095	0.010	0.012
60	-1.263	-0.031	0.008	0.010	-0.072	**0.018**	-0.045	-0.083	0.016	0.016
120	-1.100	0.031	0.049	0.023	0.038	**0.061**	0.029	0.033	0.036	0.034
12-step-ahead prediction										
3	-1.940	-0.096	0.031	**0.091**	-0.098	0.041	-0.096	-0.119	0.086	0.094
12	-1.850	-0.071	-0.012	**0.048**	-0.147	-0.004	-0.089	-0.168	0.048	0.051
36	-1.688	-0.017	-0.001	0.023	-0.122	0.019	-0.049	-0.137	0.032	**0.033**
60	-1.560	0.005	0.036	0.024	-0.068	**0.066**	-0.011	-0.073	0.034	0.036
120	-1.387	0.099	0.151	0.030	0.101	**0.193**	0.098	0.092	0.047	0.045

Notes: The first column shows the log predictive score based on the RW-C. Other columns show the difference of log predictive score from the first column. Log predictive score differences represent percentage point differences. Therefore, a difference of 0.1 corresponds to a 10% more accurate density forecast. The log predictive score from the best model is in bold letter for each variable and forecast horizon.

month ahead predictions for short term maturities when compared to most competitors.[9] By producing improved estimates of the current state of volatility, we would expect that short horizon forecasts have the largest gain. The improved density forecasting performance for the DNS-RV and RW-RV models continues even up to a 6-month forecasting horizon. At one year ahead, most models with realized volatility have their volatility processes returning close to the unconditional mean, so the gain diminishes.

Second, comparing RW-SV to RW-SV-RW, RW-RV to RW-RV-RW, and DNS-SV to DNS-SV-RW shows that given a fixed conditional mean specification, a random walk specification on conditional volatility dynamics in general leads to poorer results. This illustrates the fact that even though conditional mean dynamics of bond yields approximate a random walk in

[9]We also compute the model confidence set of Hansen, Lunde, and Nason (2011) and find the similar result. See the supporting material for the model confidence set.

Table 6 Log predictive score comparison: p-values

Maturity	RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	RW-SV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
1-step-ahead prediction									
3	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	**0.00**	**0.00**	0.13	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
36	**0.00**	0.22	**0.01**	**0.00**	0.18	0.10	0.11	**0.01**	**0.01**
60	0.08	0.93	**0.02**	0.10	0.88	0.87	0.72	0.17	0.15
120	0.07	0.99	0.94	0.11	0.82	0.93	0.98	0.30	0.22
3-step-ahead prediction									
3	**0.00**	**0.00**	0.17	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	**0.00**	**0.03**	0.93	**0.00**	**0.01**	**0.02**	**0.00**	**0.00**	**0.00**
36	0.20	0.31	0.28	0.22	0.18	0.21	0.12	0.29	0.27
60	0.53	0.60	0.27	0.59	0.37	0.51	0.34	0.67	0.66
120	0.60	0.96	0.64	0.77	0.96	0.91	0.87	0.94	0.97
6-step-ahead prediction									
3	0.12	**0.04**	0.93	0.25	**0.04**	**0.02**	**0.03**	0.24	0.32
12	0.38	0.28	0.98	0.44	0.15	0.20	0.12	0.45	0.50
36	0.57	0.16	0.74	0.64	0.19	0.09	0.16	0.73	0.75
60	0.77	0.20	0.74	0.90	0.18	0.12	0.14	0.98	0.97
120	0.35	0.52	0.78	0.55	0.64	0.51	0.59	0.68	0.66
12-step-ahead prediction									
3	0.76	0.10	0.82	0.70	0.30	0.06	0.25	0.70	0.66
12	0.97	0.23	0.80	0.38	0.41	0.11	0.36	0.34	0.35
36	0.72	**0.02**	0.17	0.92	0.43	**0.00**	0.38	0.69	0.72
60	0.10	**0.00**	0.26	0.44	0.34	**0.00**	0.32	0.51	0.56
120	**0.01**	**0.00**	0.39	0.05	0.26	**0.00**	0.23	0.07	0.07

Notes: This table presents the p-values from Amisano and Giacomini (2007) tests comparing the hypothesis of equal log predictive score of the DNS-RV with alternative models. Bold letter indicates p-values less than 5%. Test statistics are computed with a variance estimator robust to serial correlation using a rectangular kernel of $h-1$ lags and the small-sample correction proposed by Harvey, Leybourne, and Newbold (1997).

our sample, conditional volatility dynamics exhibit mean reversion. Bond yields therefore do have forecastability, although simply looking at the conditional mean dynamics do not reveal this fact strongly.

Third, an alternative specification for introducing stochastic volatility into the model by putting it on the measurement equation does not forecast as well as the specification with stochastic volatility on the transition equation. The measurement error specifications give consistent improvements in the log predictive score over and above constant volatility models, although they have a similar log predictive score pattern as DNS-C. Comparing DNS-SV to DNS-ME-SV shows that for short horizon forecasts, DNS-SV performs better whereas for longer horizon forecasts, DNS-ME-SV does better. A similar story holds when looking at DNS-RV and DNS-ME-RV, although DNS-RV does better even up to 6-month

horizon forecasts with mixed 12-month horizon results.

One fact holding back the performance of the measurement error specifications is that the measurement error variance explains a small portion of total bond yield variance. For example, the ratio of the standard deviations of smoothed measurement errors to the standard deviations of smoothed factors in the DNS-ME-SV model is often below 3 percent and never above 8 percent [10]. In figure 3, we see that the model with stochastic volatility on the measurement errors does not generate movements in the conditional time-varying volatility in various middle maturity yields. In fact, the conditional variances of the 1-year, 3-year, 5-year, and 8-year maturities are nearly on top of the black dotted line, which is the variance of the yields implied by the DNS-C model. Therefore, putting time-varying volatility in the measurement errors does not drastically change the model-implied predictive distributions. This explains why the density forecasting performance for this class of models mimics that of the DNS-C model.

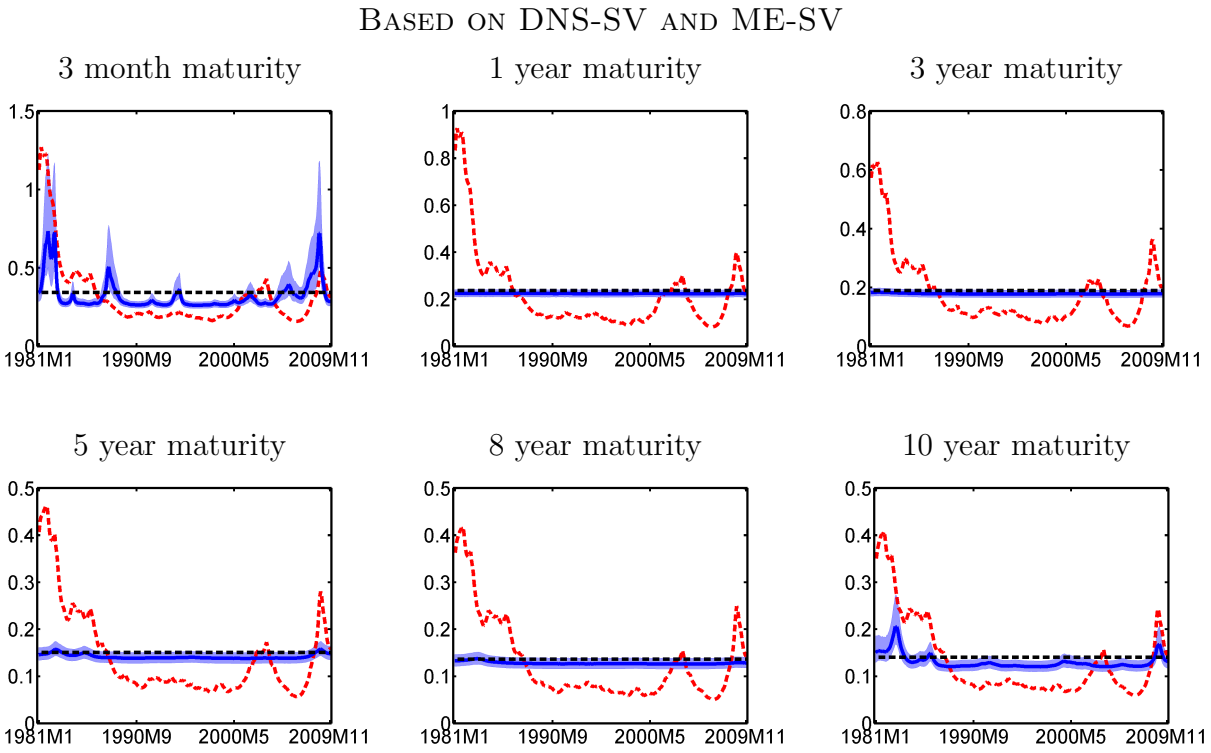
In contrast, the figure 3 shows that putting stochastic volatility in the transition equation can better capture time-varying volatility. The DNS-SV model-implied time-varying volatility consistently fits the narrative evidence of the Great Moderation from the mid-1980's until the mid-2000's. It also picks up the increases in volatility from the early 2000's recession and recent financial crises.

5.4 Forecasting empirical factors

A popular decomposition that we use as well to model the yield curve involves the level, slope, and curvature factors. The empirical counterparts of these factors are the 10-year yield for the level, 10-year minus 6-month spread for the slope, and 6-month + 10-year − 2∗5-year linear combination for the curvature. As discussed in Diebold and Rudebusch

[10]This fact is similar across different model specifications. The values from all models are presented in the supporting materials.

Figure 3 Stochastic variance of individual yield



Notes: Red dotted line: Conditional variance from DNS-SV. Blue solid line: Conditional variance from DNS-ME-SV with 80% credible interval. Black dotted horizontal line: Conditional variance from DNS-C. Estimation sample is from January 1981 to November 2009.

(2012), the level factor is related to inflation expectations while the slope factor corresponds to the business cycle.

Given that many people think about the yield curve in terms of these three factors, we also provide point and density forecasting results in this dimension. Considering this dimension provides a unique challenge for our models, as they must jointly predict the future movements of the yields correctly. This is an aspect of multivariate forecasting not present in a univariate context.

Table 7 shows the RMSE values relative to the RW-C benchmark. In the short-run, RW-RV does well for slope and curvature and DNS-RV does well for curvature. All values, however, are not significantly different from a random walk, largely replicating the result for

the individual yields. As the forecasting horizon increases, the stationary models begin to forecast well, as also pointed out in Duffee (2011). This is in line with results we find for the stationary models when forecasting the individual yields.

For the short- and long-run forecasts, allowing for time-varying volatility or adding realized volatility does not help in point prediction. Note that from table 4, adding in realized volatility largely improves the point prediction results when comparing DNS-SV to DNS-RV. This fact does not hold when forecasting the empirical factors. We believe there may be two factors at play that lead to this result. First, DNS-RV occasionally underperforms DNS-SV in forecasting the 10-year yield, an important component in the empirical factors. Second, joint prediction brings in the importance of the covariance between the yields. The root mean square error of the slope forecast, for example, depends on the RMSEs of the 6-month and 10-year yields, but also on the covariance between the errors of the two series. Specifically, a negative covariance between the forecast errors of the two series worsens RMSE. A negatively correlated forecast error could also cause the negative DNS-RV result. Perhaps by modeling the covariance between the yields more seriously and adding realized covariance data, second moment information could improve point forecasts.

For the density forecasts, allowing for stochastic volatility helps improve the joint prediction, especially in the short-run. Its effect diminishes as the horizon gets longer, similar to the individual yield results. In contrast to the individual yield prediction results, the random walk assumption on stochastic volatility no longer hurts prediction. Incorporating realized volatility information, however, worsens the density prediction of slope and curvature, although the difference is not significant for slope. A potential explanation for this result is that by only adding realized volatility information, we weight the DNS-RV estimates of h_t too strongly towards matching the individual variances of the yields, while neglecting the covariances. Adding realized covariance here could also reverse this fact.

Table 7 RMSE comparison: Empirical factors

Factor	RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
1-step-ahead prediction										
Level	0.277	1.001	**0.996**	0.988	0.989	0.999	1.000	0.988	0.986	0.997
Slope	0.295	0.995	**0.984**	1.014	1.008	1.007	0.998	1.010	1.020	0.999
Curvature	0.263	1.001	**0.968**	1.018	0.995	0.983	1.004	0.995	0.991	1.034
3-step-ahead prediction										
Level	0.489	1.000	0.998	**0.982**	0.987	0.995	1.000	0.988	0.987	0.987
Slope	0.489	0.999	**0.995**	1.023	1.018	1.027	1.001	1.018	1.023	1.020
Curvature	0.438	1.003	0.999	0.982	**0.955***	0.997	1.003	0.955*	0.974	0.996
6-step-ahead prediction										
Level	0.665	0.999	0.992*	**0.977**	0.982	0.980	0.998	0.986	0.981	0.980
Slope	0.773	1.000	0.996	**0.989**	1.012	1.034	1.000	1.012	0.990	0.991
Curvature	0.651	1.002	1.004	0.930	0.915***	0.995	1.002	**0.915*** **	0.930	0.942
12-step-ahead prediction										
Level	0.833	1.002	0.998	0.976	0.994	0.966	1.001	0.994	0.980	**0.976**
Slope	1.254	1.000	0.998**	**0.917**	0.985	1.013	1.000	0.986	0.920**	0.922**
Curvature	0.827	1.002**	1.003	0.908	0.923*	0.969	1.002***	0.922*	**0.905**	0.918

Notes: The first column shows the RMSE based on RW-C. Other columns show the relative RMSE compared to the first column. The RMSE from the best model is in bold letter. Units are in percentage points. Empirical factor is defined as: Level: 10-year yield, Slope: 10Y-6M spread, Curvature: 6M+10Y-2*5Y. Divergences in accuracy that are statistically different from zero are given by * (10%), ** (5%), *** (1%). We construct the p-values based on the Diebold and Mariano (1995) t-statistics with a variance estimator robust to serial correlation using a rectangular kernel of $h-1$ lags and the small-sample correction proposed by Harvey, Leybourne, and Newbold (1997).

6 Conclusion

We investigate the effects of introducing realized volatility information on US bond yield density forecasting. We compare the performance of our benchmark model DNS-RV with a variety of different models proposed in the literature and find that the DNS-RV model produces superior density forecasts, especially for the short-run. In addition to this, incorporating time-varying volatility in general improves density prediction, time-varying volatility is better modeled as a stationary process as opposed to a random walk, and time-varying volatility in the factor equation generates better density predictions when compared to time-varying volatility on the measurement equation.

The results for the joint forecasting performance show that there exists promising future work to be done in this area. Key in joint forecasting performance is capturing the correct time-varying covariance in the yields. Perhaps explicitly modeling correlation between the

Table 8 Log predictive score: Empirical factors

Factor	RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-RV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
1-step-ahead prediction										
Level	-0.248	0.110	0.111	0.011	**0.121**	0.110	0.106	0.108	0.048	0.037
Slope	-0.250	0.130	0.104*	-0.003	0.115	0.078	**0.134**	0.126	0.050	0.061
Curvature	-0.133	0.052**	0.008	-0.007	0.062**	0.005	0.056**	**0.065****	0.037	0.010
3-step-ahead prediction										
Level	-0.779	0.033	0.020	0.014	0.033	0.031	0.036	**0.038**	0.027	0.029
Slope	-0.744	**0.063**	0.053	-0.015	0.037	0.022	0.060	0.038	0.007	0.015
Curvature	-0.605	0.040*	0.003	0.023	0.098***	0.007	0.043**	**0.103*****	0.031	0.014
6-step-ahead prediction										
Level	-1.100	0.031	0.049	0.023	0.038	**0.061**	0.029	0.033	0.036	0.034
Slope	-1.180	-0.255	-0.064	**0.008**	-0.182	-0.101	-0.141	-0.175	0.006	0.002
Curvature	-0.995	-0.034	-0.006	0.069	0.066*	0.003	-0.031	**0.069****	0.061	0.045
12-step-ahead prediction										
Level	-1.387	0.099	0.151	0.030	0.101	**0.193**	0.098	0.092	0.047	0.045
Slope	-1.683	-0.630	-0.057	**0.044**	-0.687	-0.088	-0.289	-0.364	0.019	0.017
Curvature	-1.234	-0.062	-0.031	**0.077**	-0.034	0.021	-0.077	-0.032	0.077	0.061

Notes: For this table, we perform a one-sided test with the alternative hypothesis that the alternative model has larger log predictive score than DNS-RV. The first column shows the log predictive score based on RW-C. Other columns show the difference of log predictive score from the first column. The log predictive score from the best model is in bold letter. Empirical factor is defined as: Level: 10-year yield, Slope: 10Y-6M spread, Curvature: 6M+10Y-2*5Y. Log predictive score differences represent percentage point differences. Therefore, a difference of 0.1 corresponds to a 10% more accurate density forecast. Gains in accuracy that are statistically different from zero are given by * (10%), ** (5%), *** (1%). We construct the p-values based on the Diebold and Mariano (1995) t-statistics with a variance estimator robust to serial correlation using a rectangular kernel of $h-1$ lags and the small-sample correction proposed by Harvey, Leybourne, and Newbold (1997).

stochastic volatility innovations in conjunction with using realized covariance data could lead to large gains in joint density forecasting.

References

Amisano, G., and R. Giacomini (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business & Economic Statistics*, 25(2), 177–190.

Andersen, T. G., and L. Benzoni (2010): "Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models," *The Journal of Finance*, 65(2), 603–653.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

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 (Statistical Methodology)*, 64(2), 253–280.

Bianchi, F., H. Mumtaz, and P. Surico (2009): "The Great Moderation of the Term Structure of UK Interest Rates," *Journal of Monetary Economics*, 56(6), 856 – 871.

Carriero, A., T. E. Clark, and M. Marcellino (2013): "No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates," Discussion paper, Working paper.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541–553.

Christensen, J. H., J. A. Lopez, and G. D. Rudebusch (2011): "Can Spanned Term Structure Factors Drive Stochastic Volatility?," Discussion paper, Working paper.

Cieslak, A., and P. Povala (2013): "Understanding Bond Risk Premia," Discussion paper, Working paper.

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8(2), 262 – 302.

Del Negro, M., and F. Schorfheide (2013): "DSGE Model-Based Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann, vol. 2A, chap. 2, pp. 57 – 140. Elsevier.

Diebold, F. X., and C. Li (2006): "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130(2), 337 – 364.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 20(1), 134–144.

Diebold, F. X., and G. D. Rudebusch (2012): *Yield Curve Modeling and Forecasting*. Princeton University Press.

Duffee, G. R. (2011): "Forecasting with the Term Structure: The Role of No-Arbitrage Restrictions," Discussion paper, Working papers//the Johns Hopkins University, Department of Economics.

Duffee, G. R. (2012): "Forecasting Interest Rates," Economics Working Paper Archive 599, The Johns Hopkins University,Department of Economics.

Egorov, A. V., Y. Hong, and H. Li (2006): "Validating Forecasts of the Joint Probability Density of Bond Yields: Can Affine Models Beat Random Walk?," *Journal of Econometrics*, 135(1 – 2), 255 – 284.

Geweke, J., and G. Amisano (2010): "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, 26(2), 216 – 230.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The U.S. 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.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The Model Confidence Set," *Econometrica*, 79(2), 453–497.

Harvey, D., S. Leybourne, and P. Newbold (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13(2), 281 – 291.

Hautsch, N., and Y. Ou (2012): "Analyzing Interest Rate Risk: Stochastic Volatility in the Term Structure of Government Bond Yields," *Journal of Banking & Finance*, 36(11), 2988 – 3007.

Hautsch, N., and F. Yang (2012): "Bayesian Inference in a Stochastic Volatility Nelson–Siegel Model," *Computational Statistics & Data Analysis*, 56(11), 3774 – 3792.

Jin, X., and J. M. Maheu (2013): "Modeling Realized Covariances and Returns," *Journal of Financial Econometrics*, 11(2), 335–369.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2013): "Smooth Dynamic Factor Analysis with Application to the US Term Structure of Interest Rates," *Journal of Applied Econometrics*.

Justiniano, A., and G. E. Primiceri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *The American Economic Review*, 98(3), pp. 604–641.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), 361–393.

Koopman, S. J., M. I. P. Mallee, and M. Van der Wel (2010): "Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson–Siegel Model With Time–Varying Parameters," *Journal of Business & Economic Statistics*, 28(3), 329–343.

Maheu, J. M., and T. H. McCurdy (2011): "Do High-Frequency Measures of Volatility Improve Forecasts of Return Distributions?," *Journal of Econometrics*, 160(1), 69 – 76.

Shephard, N., and K. Sheppard (2010): "Realising the Future: Forecasting with High-

Frequency-Based Volatility (HEAVY) Models," *Journal of Applied Econometrics*, 25(2), 197–231.

Takahashi, M., Y. Omori, and T. Watanabe (2009): "Estimating Stochastic Volatility Models Using Daily Returns and Realized Volatility Simultaneously," *Computational Statistics & Data Analysis*, 53(6), 2404 – 2426.

van Dijk, D., S. J. Koopman, M. van der Wel, and J. H. Wright (2013): "Forecasting Interest Rates with Shifting Endpoints," Tinbergen Institute Discussion Papers 12-076/4, Tinbergen Institute.

Appendices

A State Space Representation

For completeness, we present the full specification of the state space form of the model. We give a detailed explanation of these equations in sections 2.1 and 2.2 of the main text. Consider a set of bond yields $y_t = \{y_t(1), ..., y_t(N)\}'$. τ_j is the maturity in months of bond yield j and λ is the point of maximal curvature.

$$y_t = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{\lambda\tau_1} \\ . & . & . \\ . & . & . \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{\lambda\tau_N} \end{pmatrix} \begin{pmatrix} l_t \\ s_t \\ c_t \end{pmatrix} + \epsilon_t \tag{A.1}$$

$$\epsilon_t \sim N(0, Q) \tag{A.2}$$

$$\log(RV_t) = \beta + \Lambda_h \tilde{h}_t + \zeta_t \tag{A.3}$$

$$\zeta_t \sim N(0, S) \tag{A.4}$$

$$\begin{pmatrix} l_t - \mu_l \\ s_t - \mu_s \\ c_t - \mu_c \end{pmatrix} = \begin{pmatrix} \phi_l & 0 & 0 \\ 0 & \phi_s & 0 \\ 0 & 0 & \phi_c \end{pmatrix} \begin{pmatrix} l_{t-1} - \mu_l \\ s_{t-1} - \mu_s \\ c_{t-1} - \mu_c \end{pmatrix} + \eta_t \tag{A.5}$$

$$\eta_t \sim N(0, H_t) \tag{A.6}$$

$$H_t = \begin{pmatrix} e^{h_{l,t}} & 0 & 0 \\ 0 & e^{h_{s,t}} & 0 \\ 0 & 0 & e^{h_{c,t}} \end{pmatrix} \tag{A.7}$$

$$h_{i,t} - \mu_{i,h} = \phi_{i,h}(h_{i,t-1} - \mu_{i,h}) + e_{i,t} \tag{A.8}$$

$$e_{i,t} \sim N(0, \sigma^2_{i,h}) \tag{A.9}$$

for $i = l, s, c$ and Q and S are diagonal matrix.

B Measurement Equation for RV: Derivation and Approximation

Equation A.3 is the linearized version of the nonlinear measurement equation that comes from adding realized volatility information to the dynamic factor model. We perform a first-order approximation of the logarithm of the following equation

$$\begin{aligned} RV_t \approx Var_{t-1}(y_t) &= diag(\Lambda_f H_t \Lambda_f' + Q) \\ &= diag(\tilde{\Lambda}_f \tilde{H}_t \tilde{\Lambda}_f' + Q) \end{aligned} \tag{A.10}$$

where we write the logarithm of volatility in deviation form $\tilde{h}_{i,t} = h_{i,t} - \mu_{h,i}$ for $i = l, s, c$. Then $\tilde{H}_t$ is a 3×3 diagonal matrix with each element corresponding to $e^{\tilde{h}_{i,t}}$ and $\tilde{\Lambda}_f = \Lambda_f[e^{\mu_l/2}, e^{\mu_s/2}, e^{\mu_c/2}]'$. We first derive the nonlinear measurement equation that links the realized volatility with underlying factor volatility. Our derivation is similar to Maheu and McCurdy (2011) but we derive it under the dynamic factor model framework. Then, we

describe the approximation to get the linearized measurement equation for RV_t.

Derivation of the measurement equation We assume that the RV_t has a log-Normal distribution (element-wise). Then we have

$$E_{t-1}[RV_t] = \exp\left(E_{t-1}\log(RV_t) + \frac{1}{2}Var_{t-1}(\log(RV_t))\right) = diag(\tilde{\Lambda}_f \tilde{H}_t \tilde{\Lambda}_f' + Q).$$

where the second equality is from Corollary 1 of Andersen, Bollerslev, Diebold, and Labys (2003) and we assume that RV_t is an unbiased estimator for the quadratic variation. Taking logarithm on both sides gives,

$$E_{t-1}\log(RV_t) + \frac{1}{2}Var_{t-1}(\log(RV_t)) = \log\left(diag(\tilde{\Lambda}_f \tilde{H}_t \tilde{\Lambda}_f' + Q)\right).$$

Assume the conditional variance of the RV_t is constant and write $E_{t-1}[\log(RV_t)] = \log(RV_t) + \zeta_t$ where[11] $\zeta_t \sim \mathcal{N}(0, S)$. Then we get

$$\log(RV_t) = \tilde{\beta} + \log(diag(\tilde{\Lambda}_f \tilde{H}_t \tilde{\Lambda}_f' + Q)) + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0, S). \tag{A.11}$$

where $\tilde{\beta}$ can be viewed as the conditional variance of log realized volatility plus potential bias caused by the assumption we made.

Linearization We present the derivation of equation 5. We linearize the equation A.11 for ith element around $\tilde{h}_{j,t} = 0$ to get

$$\log(RV_{i,t}) = \beta_i + \nu_i \left(\sum_{j=1}^{3} \tilde{\Lambda}_{f,i,j}^2 \tilde{h}_{j,t}\right) + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0, S).$$

[11] Distribution of ζ_t is obtained by assuming that RV_t follows a log-Normal distribution.

where

$$\beta_i = \tilde{\beta}_i + \log(\tilde{\Lambda}^2_{f,i} + Q_{i,i})$$
$$\nu_i = \frac{1}{\left(\tilde{\Lambda}^2_{f,i} + Q_{i,i}\right)}.$$

$\nu_i \left(\sum_{j=1}^{3} \tilde{\Lambda}^2_{f,i,j} \tilde{h}_{j,t}\right)$ corresponds to $\Lambda_{h,i}\tilde{h}_t$ and $\tilde{\Lambda}^2_{f,i,j}$ is the (i,j)th element of $\tilde{\Lambda}^2_f$ in equation 5. Note that $\beta \neq \tilde{\beta}$ as β is the constant term from the linearized equation while $\tilde{\beta}$ is from the nonlinear one.

C Estimation Procedure

Presented is the algorithm for the Gibbs sampling. We draw 15000 samples, saving every 5th draw, with the first 5000 draws as burn-in. The priors we choose for the model are all extremely loose. They are presented in table A-1.

Because of the assumption of independent AR(1) factor and stochastic volatility processes, the algorithm simplifies slightly. A general multivariate case, however, is a trivial extension.

Call $\Theta^* = \{\mu_f, \phi_f, \mu_h, \phi_h, \beta, Q, S, \sigma_h^2, f_{1:T}, h_{1:T}\}$, the parameters on which we would like to perform inference. Note that as is standard in Bayesian estimation, we use the data augmentation method and consider f_t and h_t as random vectors.

1. Initialize Θ^*.

   We do so using the Hautsch and Yang (2012) estimates where possible. For S, we initialize with the identity matrix. For each element of β, we initialize at 0. For the factors and stochastic volatilities we draw each element from a normal distribution.

   Enter into iteration i:

2. **(Drawing $Q|\Theta^*_{-Q}$)** Since Q is diagonal, we draw the diagonal elements one at a time. Each element of the diagonal term on Q is distributed as an inverse gamma distribution.[12]

3. **(Drawing $\beta, S|\Theta^*_{-\beta,S}$)** We can likewise draw β and the diagonal elements of S equation-by-equation. It is a standard linear regression normal-inverse gamma framework.

4. **(Drawing $f_{1:T}|\Theta^*_{-f_{1:T}}$)** The Carter and Kohn (1994) multi-move Gibbs sampling procedure with stochastic volatility can be used to draw the level, slope, and curvature factors.

5. **(Drawing $\mu_f, \phi_f|\Theta^*_{-\mu_f,\phi_f}$)** Because we specify the factors and stochastic volatilities to have independent AR(1) processes, we can separate the drawing of the parameters for each factor. Drawing the parameters equation-by-equation is possible through the linear regression with stochastic volatility laid out in Hautsch and Yang (2012).

6. **(Drawing $h_{1:T}|\Theta^*_{-h_{1:T}}$)** We have a measurement equation made up of two parts. The first part uses the Kim, Shephard, and Chib (1998) method to transform the level, slope, and curvature factor equations. The measurement equation defined by the level factor is

$$\log((l_t - (1-\phi_{l,h})\mu_{l,h} - \phi_{l,h} l_{t-1})^2) = h_{l,t} + \log(x^2_{l,t}) \tag{A.12}$$

   We approximate the error term using a mixture of normals as in Kim, Shephard, and Chib (1998).

   The second part is the realized volatility measurement equation.

   Therefore, equations A.13 and A.3 define the measurement equation:

$$f^*_t = \mu_h + I_{3\times 3}(h_t - \mu_h) + \log(x^2_t) \tag{A.13}$$

[12] For DNS-RV, Q enters in the realized measurement equation (equation A.3). In this case, we draw Q using the Metropolis-Hastings algorithm with a proposal distribution as an inverse gamma distribution.

The transition equation is

$$\begin{pmatrix} h_{l,t} - \mu_{l,h} \\ h_{s,t} - \mu_{s,h} \\ h_{c,t} - \mu_{c,h} \end{pmatrix} = \begin{pmatrix} \phi_{l,h} & 0 & 0 \\ 0 & \phi_{s,h} & 0 \\ 0 & 0 & \phi_{c,h} \end{pmatrix} \begin{pmatrix} h_{l,t-1} - \mu_{l,h} \\ h_{s,t-1} - \mu_{s,h} \\ h_{c,t-1} - \mu_{c,h} \end{pmatrix} + e_t \tag{A.14}$$

Because of our linear approximation of the nonlinear measurement equation, we can simply use the standard Kalman Filter along with the Carter and Kohn (1994) multi-move Gibbs simulation smoother with time-varying measurement mean and innovation volatility to draw $h_{1:T}$.

7. **(Drawing** $\mu_h, \phi_h, \sigma_h^2 | \Theta^*_{-\mu_h,\phi_h,\sigma_h^2}$**)** We use a standard linear regression normal-inverse gamma framework to draw the parameters equation-by-equation[13].

D Forecasting Procedure

Presented in equations A.15 - A.17 is the forecasting algorithm that we use. Because we are performing Bayesian analysis, we explicitly take into account the parameter uncertainty when generating our forecasts. We first draw parameters from the relevant posterior distributions (j) and then simulate 10 trajectories of data given the parameter values (k). We do so for 2000 parameter draws for a total of 20000 simulated data chains from which to compare to the realized data (Del Negro and Schorfheide (2013)). Note that for the DNS-C model, we would not have equation A.17 and the H_t would become H.

$$\hat{y}_t^{j,k} = \Lambda_f \begin{pmatrix} l_t^{j,k} \\ s_t^{j,k} \\ c_t^{j,k} \end{pmatrix} + \tilde{\epsilon}_t^{j,k}, \quad \tilde{\epsilon}_t^{j,k} \sim N\left(0, Q^j\right) \tag{A.15}$$

[13]Here relies on the specifics of the linearization for our realized volatility measurement equation. Note that we linearize that equation around the previous draw's μ_h. This means we do not have to take this equation into account when drawing our new μ_h.

$$\begin{pmatrix} l_t^{j,k} - \mu_l^j \\ s_t^{j,k} - \mu_s^j \\ c_t^{j,k} - \mu_c^j \end{pmatrix} = \begin{pmatrix} \phi_l^j & 0 & 0 \\ 0 & \phi_s^j & 0 \\ 0 & 0 & \phi_c^j \end{pmatrix} \begin{pmatrix} l_{t-1}^{j,k} - \mu_l^j \\ s_{t-1}^{j,k} - \mu_s^j \\ c_{t-1}^{j,k} - \mu_c^j \end{pmatrix} + \tilde{\eta}_t^{j,k}, \quad \tilde{\eta}_t^{j,k} \sim N\left(0, H_t^{j,k}\right) \tag{A.16}$$

$$h_{i,t}^{j,k} - \mu_{i,h}^j = \phi_{i,h}^j (h_{i,t-1}^{j,k} - \mu_{i,h}^j) + \tilde{e}_{i,t}^{j,k}, \quad \tilde{e}_{i,t}^{j,k} \sim N\left(0, \left(\sigma_{i,h}^j\right)^2\right) \tag{A.17}$$

$$j = 1, ..., 2000$$

$$k = 1, ..., 10$$

$$t = T, ..., T + 12$$

where T is the beginning of the forecasting period.

E Prior Specification

We present prior distributions in table A-1.

Table A-1 Prior Distribution

Parameter	Description	Dim.	Dist.	Para(1)	Para(2)
H	Variance of the measurement error (y_t).	17×1	IG	0	0.001
μ_f	Long-run mean parameter for f_t.	3×1	N	0	100
ϕ_f	AR(1) coefficient for f_t.	3×1	N	0.8	100
μ_h	Long-run mean parameter for h_t.	3×1	N	0	100
ϕ_h	AR(1) coefficient for h_t.	3×1	N	0.8	100
σ_h^2	Variance of the innovation for the h_t.	3×1	IG	0.01	2
β	Intercepts in the RV measurement equation. Only used for $\mathcal{M}_{RV}$	17×1	N	0	100
S	Variance of the measurement error RV. Only used for $\mathcal{M}_{RV}$	17×1	IG	0	0.001
σ_f^2	Variance of the innovation for the f_t. Only used for models without time-varying volatility	3×1	IG	0.1	2

Note: a) All prior distributions are independent. For example, prior distributions for elements in H are independent from each other and follow the inverse gamma distribution.
b) Dim: Dimension of the parameters.
c) IG: Inverse gamma distribution. Para(1) and Para(2) mean scale and shape parameters, respectively.
d) N: Normal distribution. Para(1) and Para(2) stand for mean and variance, respectively.
e) Priors for ϕ_f and ϕ_h are truncated so that the processes for factors and volatilities are stationary.
f) $\mathcal{M}_{RV}$ is the set of models with realized volatility data.

F Supporting Material (Not for publication)

F.1 Descriptive statistics of data

Table A-2 Descriptive Statistics (Yields)

Maturity	mean	std	min	max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3	5.35	3.14	0.04	16.02	0.97	0.65	0.39
6	5.52	3.17	0.15	16.48	0.98	0.66	0.40
9	5.64	3.19	0.19	16.39	0.98	0.67	0.42
12	5.75	3.19	0.25	16.10	0.98	0.69	0.44
15	5.87	3.21	0.38	16.06	0.98	0.70	0.46
18	5.95	3.20	0.44	16.22	0.98	0.71	0.47
21	6.03	3.19	0.53	16.17	0.98	0.71	0.49
24	6.06	3.15	0.53	15.81	0.98	0.72	0.50
30	6.18	3.11	0.82	15.43	0.98	0.73	0.52
36	6.29	3.08	0.98	15.54	0.98	0.74	0.54
48	6.48	3.02	1.02	15.60	0.98	0.75	0.57
60	6.60	2.94	1.56	15.13	0.98	0.76	0.59
72	6.73	2.92	1.53	15.11	0.98	0.77	0.61
84	6.81	2.84	2.18	15.02	0.98	0.77	0.61
96	6.90	2.81	2.11	15.05	0.98	0.78	0.63
108	6.95	2.79	2.15	15.11	0.98	0.78	0.63
120	6.95	2.72	2.68	15.19	0.98	0.77	0.63

Notes: For each maturity we present mean, standard deviation, minimum, maximum and the j-th order autocorrelation coefficients for j = 1, 12, and 24.

Table A-3 Descriptive Statistics (log realized volatility)

Maturity	mean	std	min	max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3	-3.05	1.29	-5.33	1.10	0.65	0.31	0.19
6	-3.35	1.21	-6.66	0.64	0.69	0.28	0.18
9	-3.24	1.10	-5.90	0.47	0.66	0.23	0.10
12	-3.11	1.02	-5.76	0.41	0.63	0.21	0.04
15	-2.99	0.96	-5.68	0.36	0.61	0.20	0.01
18	-2.89	0.93	-5.59	0.29	0.61	0.20	-0.01
21	-2.81	0.90	-5.52	0.20	0.60	0.20	-0.02
24	-2.75	0.88	-5.46	0.11	0.60	0.20	-0.03
30	-2.66	0.85	-5.34	-0.04	0.59	0.20	-0.03
36	-2.61	0.83	-5.23	-0.12	0.60	0.21	-0.03
48	-2.57	0.79	-5.02	-0.19	0.61	0.22	-0.02
60	-2.57	0.77	-4.84	-0.22	0.61	0.23	-0.01
72	-2.58	0.76	-4.67	-0.25	0.61	0.25	0.00
84	-2.59	0.76	-4.66	-0.21	0.62	0.27	0.00
96	-2.61	0.75	-4.66	-0.18	0.63	0.29	0.01
108	-2.62	0.75	-4.66	-0.17	0.64	0.30	0.01
120	-2.63	0.75	-4.65	-0.18	0.66	0.31	0.01

Notes: For each maturity we present mean, standard deviation, minimum, maximum and the j-th order autocorrelation coefficients for j = 1, 12, and 24.

F.2 In-sample estimation (posterior moments)

We denote $\{y_1, y_2, y_3, ..., y_{16}, y_{17}\}$ = monthly U.S. Treasury yields with maturities of (3, 6, 9, 12, 15, 18, 21 months and 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10 years.

Table A-4 Posterior moments of H

		RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW
y_1	5%	0.064	0.063	0.060	0.063	0.063	0.488	0.063	0.063
	50%	0.074	0.072	0.068	0.072	0.072	0.610	0.072	0.072
	95%	0.085	0.083	0.080	0.083	0.082	0.760	0.083	0.083
y_2	5%	0.008	0.008	0.010	0.008	0.008	0.157	0.008	0.008
	50%	0.010	0.009	0.012	0.009	0.009	0.202	0.009	0.009
	95%	0.012	0.011	0.015	0.011	0.011	0.284	0.011	0.011
y_3	5%	0.002	0.002	0.004	0.002	0.002	0.043	0.002	0.002
	50%	0.003	0.003	0.005	0.002	0.002	0.067	0.002	0.003
	95%	0.004	0.003	0.006	0.003	0.003	0.138	0.003	0.003
y_4	5%	0.003	0.003	0.004	0.003	0.003	0.018	0.003	0.003
	50%	0.004	0.004	0.005	0.004	0.004	0.038	0.004	0.004
	95%	0.005	0.005	0.005	0.005	0.005	0.097	0.005	0.005
y_5	5%	0.006	0.006	0.006	0.006	0.007	0.017	0.006	0.006
	50%	0.007	0.007	0.007	0.007	0.007	0.035	0.007	0.007
	95%	0.008	0.009	0.008	0.009	0.009	0.086	0.009	0.009
y_6	5%	0.006	0.006	0.005	0.006	0.006	0.016	0.006	0.006
	50%	0.006	0.006	0.006	0.006	0.007	0.034	0.007	0.007
	95%	0.007	0.008	0.007	0.007	0.008	0.083	0.008	0.008
y_7	5%	0.004	0.004	0.004	0.004	0.004	0.015	0.004	0.004
	50%	0.005	0.005	0.005	0.005	0.005	0.033	0.005	0.005
	95%	0.006	0.006	0.005	0.006	0.006	0.080	0.006	0.006
y_8	5%	0.002	0.002	0.002	0.002	0.002	0.014	0.002	0.002
	50%	0.002	0.002	0.003	0.002	0.002	0.033	0.002	0.002
	95%	0.003	0.003	0.003	0.003	0.002	0.082	0.002	0.003
y_9	5%	0.002	0.002	0.002	0.002	0.002	0.013	0.002	0.002
	50%	0.002	0.002	0.003	0.002	0.002	0.029	0.002	0.002
	95%	0.003	0.003	0.003	0.003	0.003	0.075	0.003	0.003
y_{10}	5%	0.002	0.002	0.003	0.002	0.002	0.011	0.002	0.002
	50%	0.003	0.003	0.003	0.003	0.003	0.026	0.003	0.003
	95%	0.003	0.003	0.003	0.003	0.003	0.068	0.003	0.003
y_{11}	5%	0.004	0.004	0.004	0.004	0.004	0.009	0.004	0.004
	50%	0.004	0.004	0.004	0.004	0.004	0.021	0.004	0.004
	95%	0.005	0.005	0.005	0.005	0.005	0.059	0.005	0.005
y_{12}	5%	0.003	0.003	0.004	0.003	0.003	0.007	0.003	0.003
	50%	0.004	0.004	0.004	0.004	0.004	0.017	0.004	0.004
	95%	0.004	0.004	0.005	0.004	0.004	0.049	0.004	0.004
y_{13}	5%	0.004	0.004	0.005	0.004	0.004	0.006	0.004	0.004
	50%	0.005	0.005	0.006	0.005	0.005	0.013	0.005	0.005
	95%	0.005	0.005	0.006	0.005	0.005	0.038	0.005	0.005
y_{14}	5%	0.003	0.003	0.004	0.003	0.003	0.004	0.003	0.003
	50%	0.004	0.004	0.005	0.004	0.004	0.011	0.004	0.004
	95%	0.005	0.004	0.006	0.004	0.004	0.032	0.004	0.004
y_{15}	5%	0.003	0.002	0.004	0.003	0.002	0.004	0.002	0.002
	50%	0.003	0.003	0.005	0.003	0.003	0.011	0.003	0.003
	95%	0.004	0.004	0.006	0.004	0.004	0.031	0.004	0.004
y_{16}	5%	0.006	0.007	0.007	0.006	0.007	0.006	0.007	0.007
	50%	0.007	0.008	0.008	0.008	0.008	0.012	0.008	0.008
	95%	0.009	0.009	0.010	0.009	0.009	0.028	0.009	0.009
y_{17}	5%	0.012	0.012	0.012	0.012	0.013	0.016	0.013	0.012
	50%	0.014	0.014	0.014	0.014	0.014	0.029	0.014	0.014
	95%	0.016	0.016	0.016	0.016	0.017	0.053	0.017	0.017

Notes: Variance of the measurement error. Not applicable to DNS-ME-SV and DNS-ME-RV. Estimation sample is from January 1981 to November 2009.

Table A-5 Posterior moments of parameters related to f_t

		RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW	DNS-ME-SV	DNS-ME-RV
$\mu_{f,l}$	5%	0.00	0.00	0.00	-7.28	-5.83	-17.48	0.00	-5.44	-7.20	-7.40
	50%	0.00	0.00	0.00	4.63	4.57	-5.47	0.00	4.71	4.41	4.37
	95%	0.00	0.00	0.00	7.86	7.17	2.15	0.00	6.94	7.83	7.79
$\mu_{f,s}$	5%	0.00	0.00	0.00	-4.28	-4.04	-1.14	0.00	-4.28	-4.47	-4.49
	50%	0.00	0.00	0.00	-2.54	-0.88	2.92	0.00	-0.98	-2.55	-2.59
	95%	0.00	0.00	0.00	-1.43	6.93	13.81	0.00	6.46	-1.50	-1.43
$\mu_{f,c}$	5%	0.00	0.00	0.00	-2.64	-2.13	-6.31	0.00	-1.78	-2.97	-2.91
	50%	0.00	0.00	0.00	-0.82	-0.52	0.16	0.00	-0.42	-0.87	-0.79
	95%	0.00	0.00	0.00	0.92	0.84	8.66	0.00	0.70	1.26	1.33
$\phi_{f,l}$	5%	1.00	1.00	1.00	0.98	0.98	0.99	1.00	0.98	0.98	0.98
	50%	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99	0.99
	95%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\phi_{f,s}$	5%	1.00	1.00	1.00	0.94	0.97	0.99	1.00	0.97	0.94	0.94
	50%	1.00	1.00	1.00	0.96	0.99	1.00	1.00	0.99	0.96	0.97
	95%	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99	0.99
$\phi_{f,c}$	5%	1.00	1.00	1.00	0.92	0.91	0.97	1.00	0.90	0.92	0.93
	50%	1.00	1.00	1.00	0.95	0.95	0.99	1.00	0.94	0.96	0.96
	95%	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.98	0.99	0.99
$\sigma_{f,l}$	5%	0.10			0.10	0.98	0.99		0.98	0.09	0.09
	50%	0.12			0.11	0.99	1.00		0.99	0.11	0.10
	95%	0.13			0.13	1.00	1.00		1.00	0.12	0.12
$\sigma_{f,s}$	5%	0.16			0.16					0.16	0.15
	50%	0.19			0.19					0.18	0.17
	95%	0.22			0.21					0.20	0.19
$\sigma_{f,c}$	5%	0.42			0.43					0.40	0.37
	50%	0.49			0.51					0.46	0.44
	95%	0.57			0.59					0.54	0.51

Notes: Estimation sample is from January 1981 to November 2009.

Table A-6 Posterior moments of parameters related to h_t

		RW-C	RW-SV	RW-RV	DNS-C	DNS-SV	DNS-RV	RW-SV-RW	DNS-SV-RW
$\mu_{h,l}$	5%		-4.75	-4.11		-4.40	-4.19	0.00	0.00
	50%		-2.55	-3.85		-2.56	-3.92	0.00	0.00
	95%		-1.11	-1.76		-1.48	-2.21	0.00	0.00
$\mu_{h,s}$	5%		-3.39	-3.22		-3.41	-3.22	0.00	0.00
	50%		-2.30	-2.82		-2.30	-2.85	0.00	0.00
	95%		-1.69	-1.74		-1.66	-1.94	0.00	0.00
$\mu_{h,c}$	5%		-1.47	-2.25		-1.48	-2.18	0.00	0.00
	50%		-0.97	-1.93		-0.96	-1.88	0.00	0.00
	95%		-0.53	-1.45		-0.47	-1.39	0.00	0.00
$\phi_{h,l}$	5%		0.93	0.59		0.93	0.58	1.00	1.00
	50%		0.98	0.66		0.98	0.66	1.00	1.00
	95%		1.00	0.73		1.00	0.73	1.00	1.00
$\phi_{h,s}$	5%		0.92	0.56		0.92	0.57	1.00	1.00
	50%		0.97	0.65		0.96	0.64	1.00	1.00
	95%		0.99	0.72		1.00	0.72	1.00	1.00
$\phi_{h,c}$	5%		0.74	0.40		0.81	0.39	1.00	1.00
	50%		0.91	0.50		0.92	0.49	1.00	1.00
	95%		0.98	0.61		0.98	0.60	1.00	1.00
$\sigma_{h,l}$	5%		0.01	0.39		0.01	0.41	0.01	0.01
	50%		0.03	0.72		0.03	0.75	0.01	0.01
	95%		0.08	0.88		0.08	0.92	0.03	0.03
$\sigma_{h,s}$	5%		0.01	1.29		0.01	1.27	0.01	0.01
	50%		0.03	1.75		0.04	1.72	0.02	0.03
	95%		0.10	3.00		0.10	2.64	0.06	0.05
$\sigma_{h,c}$	5%		0.01	1.41		0.02	1.30	0.01	0.01
	50%		0.12	2.00		0.09	1.78	0.03	0.03
	95%		0.40	4.00		0.27	4.30	0.09	0.09

Notes: Parameters related to h_t. Not applicable to DNS-C, RW-C, DNS-ME-SV and DNS-ME-RV. Estimation sample is from January 1981 to November 2009.

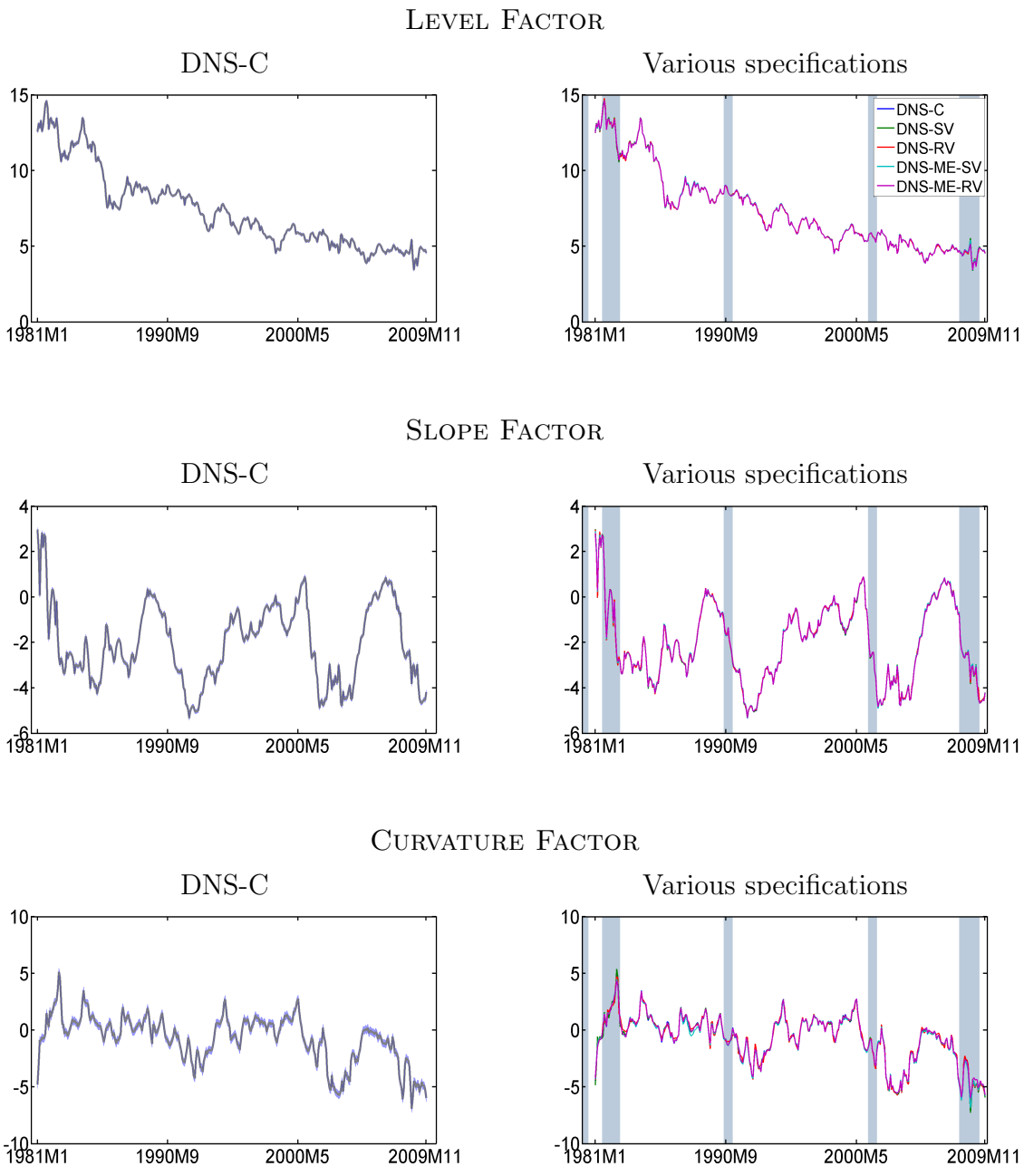
Table A-7 Posterior moments of parameters related to h_t for ME-SV model

	μ_h			ϕ_h			σ_h^2		
		DNS-ME-SV	DNS-ME-RV		DNS-ME-SV	DNS-ME-RV		DNS-ME-SV	DNS-ME-RV
y_1	5%	-4.80	-4.28	5%	0.91	0.91	5%	0.14	0.04
	50%	-3.61	-3.45	50%	0.95	0.95	50%	0.23	0.08
	95%	-2.50	-2.92	95%	0.99	0.98	95%	0.36	0.13
y_2	5%	-5.75	-5.79	5%	0.89	0.91	5%	0.03	0.01
	50%	-4.88	-5.03	50%	0.95	0.95	50%	0.07	0.03
	95%	-4.14	-4.62	95%	0.99	0.99	95%	0.14	0.06
y_3	5%	-6.28	-5.77	5%	0.53	0.88	5%	0.00	0.00
	50%	-5.48	-5.55	50%	0.96	0.93	50%	0.01	0.01
	95%	-4.96	-5.35	95%	1.00	0.98	95%	0.04	0.02
y_4	5%	-6.52	-5.60	5%	0.24	0.87	5%	0.00	0.00
	50%	-5.43	-5.43	50%	0.96	0.92	50%	0.01	0.01
	95%	-4.65	-5.26	95%	1.00	0.97	95%	0.02	0.01
y_5	5%	-7.89	-5.35	5%	0.95	0.86	5%	0.01	0.00
	50%	-5.40	-5.12	50%	0.98	0.92	50%	0.02	0.01
	95%	-4.15	-4.93	95%	1.00	0.97	95%	0.04	0.02
y_6	5%	-10.34	-5.69	5%	0.97	0.88	5%	0.00	0.01
	50%	-5.95	-5.39	50%	0.99	0.93	50%	0.01	0.02
	95%	-5.14	-5.14	95%	1.00	0.97	95%	0.02	0.04
y_7	5%	-9.71	-6.07	5%	0.97	0.89	5%	0.00	0.01
	50%	-6.16	-5.67	50%	0.99	0.94	50%	0.01	0.03
	95%	-5.46	-5.37	95%	1.00	0.98	95%	0.01	0.06
y_8	5%	-8.41	-6.01	5%	0.92	0.84	5%	0.00	0.00
	50%	-5.92	-5.85	50%	0.98	0.91	50%	0.01	0.01
	95%	-5.03	-5.68	95%	1.00	0.96	95%	0.01	0.02
y_9	5%	-7.06	-6.29	5%	0.93	0.87	5%	0.00	0.01
	50%	-6.11	-5.98	50%	0.97	0.93	50%	0.01	0.03
	95%	-5.63	-5.73	95%	1.00	0.97	95%	0.03	0.06
y_{10}	5%	-8.33	-5.89	5%	0.95	0.83	5%	0.00	0.00
	50%	-5.91	-5.71	50%	0.98	0.90	50%	0.01	0.01
	95%	-5.00	-5.56	95%	1.00	0.95	95%	0.01	0.02
y_{11}	5%	-7.80	-5.88	5%	0.95	0.86	5%	0.01	0.02
	50%	-5.53	-5.55	50%	0.98	0.92	50%	0.02	0.05
	95%	-0.82	-5.27	95%	1.00	0.96	95%	0.05	0.09
y_{12}	5%	-8.28	-5.68	5%	0.95	0.84	5%	0.00	0.01
	50%	-5.69	-5.50	50%	0.98	0.91	50%	0.01	0.01
	95%	-4.91	-5.31	95%	1.00	0.96	95%	0.02	0.04
y_{13}	5%	-8.00	-6.12	5%	0.95	0.88	5%	0.01	0.03
	50%	-5.48	-5.60	50%	0.99	0.94	50%	0.03	0.06
	95%	0.85	-5.20	95%	1.00	0.98	95%	0.06	0.10
y_{14}	5%	-9.18	-6.51	5%	0.97	0.92	5%	0.01	0.03
	50%	-5.94	-5.77	50%	0.99	0.96	50%	0.02	0.05
	95%	-2.26	-5.14	95%	1.00	0.99	95%	0.04	0.08
y_{15}	5%	-8.72	-6.07	5%	0.96	0.88	5%	0.00	0.01
	50%	-5.95	-5.69	50%	0.99	0.93	50%	0.01	0.03
	95%	-2.15	-5.36	95%	1.00	0.98	95%	0.03	0.07
y_{16}	5%	-9.02	-6.04	5%	0.96	0.89	5%	0.01	0.03
	50%	-5.84	-5.42	50%	0.99	0.94	50%	0.01	0.06
	95%	-4.57	-4.92	95%	1.00	0.98	95%	0.04	0.11
y_{17}	5%	-6.53	-5.66	5%	0.93	0.91	5%	0.04	0.04
	50%	-5.02	-4.87	50%	0.97	0.95	50%	0.07	0.07
	95%	-3.06	-4.25	95%	1.00	0.99	95%	0.14	0.12

Notes: Parameters related to h_t. For DNS-ME-SV and DNS-ME-RV. Estimation sample is from January 1981 to November 2009.

F.3 Extracted factors (f_t)

Figure A-1 Extracted factors



Notes: Left columns: Factors estimated from the DNS-C model with 80% credible intervals. Right column: Estimated factors from the various specifications. Shaded bars on the right panel are NBER recession dates. 1) Factors are very similar to each other. 2) Factors are very accurately estimated. Estimation sample is from January 1981 to November 2009.

F.4 Relative importance (ratio in %) of variation between the measurement error and f_t.

Table A-8 Relative importance (ratio in %) of variation between the measurement error and f_t.

Maturity	DNS-C	DNS-SV	DNS-RV	DNS-ME-SV	DNS-ME-RV
3	7.72	7.73	7.77	7.95	7.66
6	2.54	2.54	2.80	2.78	2.56
9	1.13	1.12	1.60	1.55	1.51
12	1.80	1.81	1.90	1.73	1.85
15	2.24	2.26	2.22	2.13	2.09
18	2.05	2.06	2.04	2.00	1.94
21	1.77	1.78	1.75	1.75	1.74
24	1.29	1.29	1.35	1.36	1.34
30	1.31	1.28	1.44	1.43	1.41
36	1.43	1.42	1.52	1.39	1.42
48	2.04	2.04	2.11	2.08	2.10
60	1.70	1.71	1.85	1.72	1.69
72	2.18	2.18	2.43	2.32	2.25
84	1.96	1.95	2.24	2.13	2.19
96	1.48	1.45	2.01	1.51	1.64
108	2.74	2.78	2.92	2.57	2.66
120	4.04	4.07	4.11	3.92	4.16

Notes: We calculate $\frac{std(e_t)}{std(f_t)} * 100$ where e_t is measurement error and f_t is a vector level, slope, and curvature factor. This table is to show that variation in the measurement equation is relatively smaller than variation in the factor component. Sizes of variation from the measurement error is about $1\% \sim 8$ % of the variation from the factors. Mostly they are below 3% except 3 month and 10 year bond yields. This evidence supports that time-varying volatility in the transition equation (factor equation) plays much larger role in prediction. Calculated at the posterior median. Estimation sample is from January 1981 to November 2009.

F.5 Model confidence set

Table A-9 Model Confidence Set (5%) Based on the log predictive score

Maturity	List of Models									
1-step-ahead prediction										
3	DNS-RV									
12	RW-RV	DNS-RV								
36	DNS-RV									
60	DNS-SV	DNS-RV	RW-SV							
120	DNS-RV	RW-SV	RW-RV	DNS-SV						
3-step-ahead prediction										
3	RW-RV	DNS-RV								
12	DNS-SV	RW-SV-RW	RW-SV	RW-RV	DNS-RV					
36	RW-C	DNS-SV	DNS-ME-RV	DNS-ME-SV	RW-SV-RW	RW-SV	RW-RV	DNS-RV		
60	DNS-SV-RW	DNS-SV	RW-C	DNS-C	RW-SV-RW	DNS-ME-RV	DNS-ME-SV	RW-RV	RW-SV	DNS-RV
120	RW-RV	DNS-ME-SV	DNS-ME-RV	RW-SV	DNS-RV	DNS-SV	RW-SV-RW	DNS-SV-RW		
6-step-ahead prediction										
3	DNS-SV	DNS-ME-SV	RW-SV	DNS-ME-RV	RW-RV	DNS-RV				
12	DNS-SV-RW	DNS-SV	DNS-C	RW-C	RW-SV-RW	DNS-ME-SV	RW-SV	DNS-ME-RV	RW-RV	DNS-RV
36	DNS-SV-RW	DNS-SV	RW-SV-RW	RW-SV	DNS-C	RW-C	DNS-ME-SV	DNS-ME-RV	RW-RV	DNS-RV
60	DNS-SV-RW	RW-SV-RW	DNS-SV	RW-SV	RW-C	RW-RV	DNS-C	DNS-ME-RV	DNS-ME-SV	DNS-RV
120	RW-SV-RW	DNS-C	RW-SV	DNS-SV-RW	DNS-ME-RV	DNS-ME-SV	DNS-SV	RW-RV	DNS-RV	
12-step-ahead prediction										
3	RW-RV	DNS-RV	DNS-ME-SV	DNS-C	DNS-ME-RV					
12	DNS-SV-RW	DNS-SV	RW-SV-RW	RW-SV	RW-RV	RW-C	DNS-RV	DNS-C	DNS-ME-SV	DNS-ME-RV
36	DNS-SV-RW	DNS-SV	RW-SV-RW	RW-SV	RW-RV	RW-C	DNS-C	DNS-RV	DNS-ME-SV	DNS-ME-RV
60	DNS-ME-SV	RW-RV	DNS-ME-RV	DNS-RV						
120	DNS-ME-RV	DNS-ME-SV	RW-SV	RW-SV-RW	DNS-SV-RW	DNS-SV	RW-RV	DNS-RV		

Notes: This table lists a subset of forecasting models that includes the best models (in terms of the log predictive score) at the 5% confidence level. Specifically, we define the difference in the log predictive score for model i and j as

$$d_{ij,t} = LPS_{i,t} - LPS_{j,t}$$

and define $\mu_{ij} = E[d_{ij,t}]$. Then, the set of best forecasts is defined as,

$$\mathcal{M}^* = \{i \in \mathcal{M} : \mu_{ij} \geq 0, \quad \forall j \in \mathcal{M}\}.$$

We follow Hansen, Lunde, and Nason (2011) to construct the model confidence set. We construct p-values using the stationary bootstrap with 10,000 replications and the average window length 12. Computation is based on the MFE Toolbox provided by Kevin Sheppard.