# Internet Appendix for: Term structure determinants of time-varying risk of one-year bond returns

Revansiddha Basavaraj Khanapure*
University of Texas at Dallas

*Corresponding author: Jindal School of Management, University of Texas at Dallas, 800 W Campbell Rd, JSOM 14.218, Richardson, TX 75080. Phone: +1-972-883-5874. Email: khanapure@gmail.com and rbk160130@utdallas.edu. ORCID: 0000-0002-1287-4835.
I thank Prof. John Cochrane and Prof. Lars Hansen for numerous discussions and comments. I also thank Professors Chinmay Jain, Monika Piazzesi, Jeffrey R. Russell, Ruey S. Tsay, Pietro Veronesi, participants at Financial Management Association International 2014 meetings, and seminar participants at the University of Chicago Booth School of Business for valuable comments. I thank Alexi Savov for numerous proof reads.

The following is Internet Appendix for article titled "Term structure determinants of time-varying risk of one year bond returns."

# Appendix A

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)} \;\; \forall\, n = 2, \ldots, 5 \tag{S}$$

$$\epsilon_{t+1}^{(n)} = \sqrt{h_{t+1}^{(n)}} \cdot u_{t+1}^{(n)} \;\; with \;\; u_{t+1}^{(n)} \sim N(0,1)$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)}) \tag{S-1}$$

The maximum likelihood estimation procedure uses overlapping monthly data to model the excess returns at 1-year horizon. This results in serially correlated unexpected returns $\epsilon_{t+1}^{(n)}$ in the specification pair (S), (S-1). I account for the serial correlation in three different ways. The simplest technique is just to use the yearly data instead of the monthly data. The second technique is to explicitly model the correlation with a parsimonious model. The third technique is to use monthly data and correct the standard errors to account for the correlation of the unexpected returns. I explain the second and the third procedure below. Both the procedures yield covariance matrices for the coefficient estimates that are similar but not the same. However the inference and the conclusion do not differ. I use the specification pair (S), (S-1) as an example in the following illustration. I apply the same procedures for other specification pairs.

## A.1 Dynamic Conditional Correlation Model

Dynamic Conditional Correlation (DCC) model (Engle, 2002) is a parsimonious way to account for the conditional correlations of the unexpected shocks. I separate the monthly data into 12 different annual series corresponding to each of the 12 calendar months (e.g.

Returns from January each year to January next year along with the forward rates rates in January each year form the 1$^{st}$ series). The DCC model takes into account the conditional correlations between the unexpected returns for each of the annual series (corr$(\epsilon^{(n)}_{t+1,i}, \epsilon^{(n)}_{t+1,j})$ for $i \neq j,\ i, j = \{\text{Jan}, \ldots, \text{Dec}\}$). To be specific, I fit DCC(1,1) model. The estimates of the correlation process parameters are $[0.02, 0.02]$ with the t-ratios $[0.3, 0.1]$ for $n = 2$ for the specification pair (S), (S-1). Similar values are obtained for the excess returns of bonds of maturities $n = 3, 4$ and 5. In fact a constant conditional correlation hypothesis is not rejected. The implied conditional correlations do not vary much over the sample period. The mean of the conditional correlations are in Table 14.

Mean Conditional Correlations

| | **Jan** | **Feb** | **Mar** | **Apr** | **May** | **Jun** | **Jul** | **Aug** | **Spt** | **Oct** | **Nov** | **Dec** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Jan** | 1.00 | 0.91 | 0.82 | 0.80 | 0.66 | 0.63 | 0.56 | 0.55 | 0.55 | 0.55 | 0.28 | 0.27 |
| **Feb** | 0.91 | 1.00 | 0.91 | 0.83 | 0.68 | 0.69 | 0.54 | 0.56 | 0.59 | 0.57 | 0.31 | 0.29 |
| **Mar** | 0.82 | 0.91 | 1.00 | 0.77 | 0.52 | 0.53 | 0.37 | 0.49 | 0.54 | 0.56 | 0.40 | 0.36 |
| **Apr** | 0.80 | 0.83 | 0.77 | 1.00 | 0.88 | 0.84 | 0.73 | 0.73 | 0.75 | 0.70 | 0.41 | 0.43 |
| **May** | 0.66 | 0.68 | 0.52 | 0.88 | 1.00 | 0.90 | 0.83 | 0.77 | 0.75 | 0.59 | 0.31 | 0.37 |
| **Jun** | 0.63 | 0.69 | 0.53 | 0.84 | 0.90 | 1.00 | 0.87 | 0.82 | 0.84 | 0.62 | 0.45 | 0.46 |
| **Jul** | 0.56 | 0.54 | 0.37 | 0.73 | 0.83 | 0.87 | 1.00 | 0.89 | 0.82 | 0.67 | 0.47 | 0.46 |
| **Aug** | 0.55 | 0.56 | 0.49 | 0.73 | 0.77 | 0.82 | 0.89 | 1.00 | 0.92 | 0.83 | 0.67 | 0.66 |
| **Spt** | 0.55 | 0.59 | 0.54 | 0.75 | 0.75 | 0.84 | 0.82 | 0.92 | 1.00 | 0.84 | 0.69 | 0.72 |
| **Oct** | 0.55 | 0.57 | 0.56 | 0.70 | 0.59 | 0.62 | 0.67 | 0.83 | 0.84 | 1.00 | 0.83 | 0.83 |
| **Nov** | 0.28 | 0.31 | 0.40 | 0.41 | 0.31 | 0.45 | 0.47 | 0.67 | 0.69 | 0.83 | 1.00 | 0.88 |
| **Dec** | 0.27 | 0.29 | 0.36 | 0.43 | 0.37 | 0.46 | 0.46 | 0.66 | 0.72 | 0.83 | 0.88 | 1.00 |

Table 14: Mean of the conditional correlations between the unexpected returns corr$(\epsilon^{(n)}_{t+1,i}, \epsilon^{(n)}_{t+1,j})$ for $i, j = \{\text{Jan}, \ldots, \text{Dec}\}$, $n = 2$, implied by the DCC(1,1) model for the specification pair (S), (S-1).

The non-negativity constraints (and sum less than one) for the parameters in the correlation process are not binding. I follow Engle (2002) and calculate the covariance matrix of the estimates of the parameters in the conditional mean and the conditional volatility specifications. The model explicitly accounts for the conditional correlations between the unexpected returns for each of the maturities. The standard errors implied by this model are similar but different from a model that simply considers one annual series. However the conclusions and the results remain unchanged.

A short summary of the DCC(1,1) model is included below for the specification pair (S), (S-1). For the DCC(1,1) model, I maximize the sum of the log-likelihoods over all the annual series (Jan,...,Dec) and impose the restriction that the parameters in each of the conditional mean and conditional volatility specifications are the same across the twelve different annual series. The specification in the elaborate form is,

$$\begin{aligned} rx^{(n)}_{t+1,i} &= \kappa^{(n),i} + \gamma^{(n),i}_{1,\ldots,5} \cdot f^{(1,\ldots,5)}_{t,i} + \epsilon^{(n)}_{t+1,i} \;\; \forall\, n = 2,\ldots,5; \;\; i = \text{Jan},\ldots,\text{Dec} \\ \epsilon^{(n),i}_{t+1} &= \sqrt{h^{(n)}_{t+1,i}} \cdot u^{(n)}_{t+1,i} \;\; with \;\; u^{(n)}_{t+1,i} \sim N(0,1) \\ \ln(h^{(n)}_{t+1,i}) &= \alpha^{(n),i} + \delta^{(n),i}_1 \cdot \ln((\epsilon^{(n)}_{t,i})^2 + \lambda) + \beta^{(n),i}_{1,\ldots,5} \cdot f^{(1,\ldots,5)}_{t,i}. \end{aligned}$$

Along with the restrictions,

$$\begin{aligned} \gamma^{(n),\text{Jan}}_{1,\ldots,5} &=, \ldots, = \gamma^{(n),\text{Dec}}_{1,\ldots,5}, \quad \beta^{(n),\text{Jan}}_{1,\ldots,5} =, \ldots, = \beta^{(n),\text{Dec}}_{1,\ldots,5} \\ \kappa^{(n),\text{Jan}}_{1,\ldots,5} &=, \ldots, = \kappa^{(n),\text{Dec}}_{1,\ldots,5}, \quad \alpha^{(n),\text{Jan}}_{1,\ldots,5} =, \ldots, = \alpha^{(n),\text{Dec}}_{1,\ldots,5}. \end{aligned}$$

In short the specification with the restrictions is,

$$\begin{aligned} rx^{(n)}_{t+1,i} &= \kappa^{(n)} + \gamma^{(n)}_{1,\ldots,5} \cdot f^{(1,\ldots,5)}_{t,i} + \epsilon^{(n)}_{t+1,i} \;\; \forall\, n = 2,\ldots,5; \;\; i = \text{Jan},\ldots,\text{Dec} \\ \epsilon^{(n)}_{t+1,i} &= \sqrt{h^{(n)}_{t+1,i}} \cdot u^{(n)}_{t+1,i} \;\; with \;\; u^{(n)}_{t+1,i} \sim N(0,1) \\ \ln(h^{(n)}_{t+1,i}) &= \alpha^{(n)} + \delta^{(n)}_1 \cdot \ln((\epsilon^{(n)}_{t,i})^2 + \lambda) + \beta^{(n)}_{1,\ldots,5} \cdot f^{(1,\ldots,5)}_{t,i}. \end{aligned}$$

Let, vector $\Omega$ represent the parameters in the conditional mean and the conditional volatility specification.

$$\Omega = [\kappa^{(n)}, \gamma^{(n)}_{1,\ldots,5}, \alpha^{(n)}, \delta^{(n)}, \beta^{(n)}_{1,\ldots,5}]^{\text{T}}$$

The conditional correlations between the unexpected returns $\epsilon^{(n)}_{t+1,i}$, $\epsilon^{(n)}_{t+1,j}$ is summarized in a conditional correlation matrix $R_{t+1}$. DCC(1,1) model assumes a simple evolution process for the correlation matrix $R_{t+1}$. Let $h^{(n)}_{t+1,(i,j)}$ be the covariance of the unexpected returns between two different series $i$ and $j$. The distribution assumed for the unexpected excess returns $\epsilon^{(n)}_{t+1,i}$, the evolution process for $R_{t+1}$ and the log likelihood functions are as follows.

$$\begin{aligned}
\epsilon^{(n)}_{t+1}|\Im_t &\sim N(0, H_{t+1}) \quad \text{Where } \epsilon_{t+1} = \left[\epsilon^{(n)}_{t+1,\text{Jan}}, \ldots, \epsilon^{(n)}_{t+1,\text{Dec}}\right]^{\mathrm{T}} \\
H_{t+1} &= D_{t+1}R_{t+1}D_{t+1} \\
D_{t+1} &= \text{diag}\left[(h^{(n)}_{t+1,\text{Jan}})^{1/2}, \ldots, (h^{(n)}_{t+1,\text{Dec}})^{1/2}\right] \\
R_{t+1} &= [\rho_{i,j}]_{t+1} \quad \text{with } (\rho_{ii})_{t+1} = 1 \\
\text{Thus } \; h^{(n)}_{t+1,(i,j)} &= (\rho_{i,j})_{t+1}\sqrt{h^{(n)}_{t+1,i}h^{(n)}_{t+1,j}} = E_t[u^{(n)}_{t+1,i}u^{(n)}_{t+1,j}]\sqrt{h^{(n)}_{t+1,i}h^{(n)}_{t+1,j}} \\
&= E_t[\epsilon^{(n)}_{t+1,\text{Jan}}\epsilon^{(n)}_{t+1,\text{Dec}}] \\
\epsilon^{(n)}_{t+1} &= D^{-1}_{t+1}u_{t+1} \quad \text{Where } u^{(n)}_{t+1} = \left[u^{(n)}_{t+1,\text{Jan}}, \ldots, u^{(n)}_{t+1,\text{Dec}}\right]^{\mathrm{T}}
\end{aligned}$$

The process for the correlation matrix $R_{t+1}$ is,

$$\begin{aligned}
Q_{t+1} &= (1-\theta_1-\theta_2)\bar{Q} + \theta_1 u_t \cdot u_t^{\mathrm{T}} + \theta_2 Q_t && \text{(D-1)} \\
R_{t+1} &= (\text{diag}(Q_t))^{-1/2}Q_t(\text{diag}(Q_t))^{-1/2}.
\end{aligned}$$

The restrictions are $\theta_1 > 0$, $\theta_2 > 0$ and $\theta_1 + \theta_2 < 1$. A test of $\theta_1 + \theta_2 = 0$ tests the hypothesis of a constant conditional correlation structure. Let, $\Theta = [\theta_1, \theta_2]^{\mathrm{T}}$ represent the parameters for the process for the correlation matrix. The log-likelihood function $L = L(\Omega, \Theta)$ is,

$$
\begin{aligned}
-2L &= \sum_{t=1}^{T} \left( m \ln(2\pi) + \ln|H_t| + \epsilon_{t+1}^{(n)\,\prime} H_t^{-1} \epsilon_{t+1}^{(n)} \right) \quad m = 12 \\
&= \sum_{t=1}^{T} \left( m \ln(2\pi) + \ln|D_t R_t D_t| + \epsilon_{t+1}^{(n)\,\prime} D_t^{-1} R_t^{-1} D_t^{-1} \epsilon_{t+1}^{(n)} \right) \\
&= \sum_{t=1}^{T} \left( m \ln(2\pi) + 2\ln|D_t| + \ln|R_t| + u_{t+1}^{(n)\,\prime} R_t^{-1} u_{t+1}^{(n)} \right) \\
&= \sum_{t=1}^{T} \left( m \ln(2\pi) + 2\ln|D_t| + \epsilon_{t+1}^{(n)\,\prime} D_t^{-1} D_t^{-1} \epsilon_{t+1}^{(n)} - u_{t+1}^{(n)\,\prime} u_{t+1}^{(n)} + \ln|R_t| + u_{t+1}^{(n)\,\prime} R_t^{-1} u_{t+1}^{(n)} \right) \\
&= \sum_{t=1}^{T} \left( m \ln(2\pi) + \ln|D_t|^2 + \epsilon_{t+1}^{(n)\,\prime} D_t^{-2} \epsilon_{t+1}^{(n)} \right) \\
&\quad + \sum_{t=1}^{T} \left( \ln|R_t| + u_{t+1}^{(n)\,\prime} R_t^{-1} u_{t+1}^{(n)} - u_{t+1}^{(n)\,\prime} u_{t+1}^{(n)} \right) \\
&= -2 \left( L_1(\Omega) + L_2(\Omega, \Theta) \right)
\end{aligned}
$$

In the first step I maximize $L(\Omega)$ and obtain, $\hat{\Omega}$, the estimates of the parameters in the conditional mean and the conditional volatility specification. In the second step I maximize $L(\hat{\Omega}, \Theta)$ and obtain, $\hat{\Theta}$, the estimates of the parameters in the correlation process, (D-1). I keep the coefficient estimates $\hat{\Omega}$ obtained in the first step fixed while estimating $\hat{\Theta}$ in the second step. The log-likelihood for the first maximization step is,

$$
L_1(\Omega) = -\frac{1}{2} \sum_{i=\text{Jan}}^{\text{Dec}} \sum_{t=1}^{T} \left( \ln(2\pi) + \ln(h_{t,i}^{(n)}) + \frac{(\epsilon_{t,i}^{(n)})^2}{h_{t,i}^{(n)}} \right).
$$

The log-likelihood for the second maximization step is,

$$
L_2(\hat{\Omega}, \Theta) = \sum_{t=1}^{T} \left( \ln|R_t| + u_{t+1}^{(n)\,\prime} R_t^{-1} u_{t+1}^{(n)} - u_{t+1}^{(n)\,\prime} u_{t+1}^{(n)} \right) \Big|_{\Omega=\hat{\Omega}}.
$$

## A.2 Newey-West type correction to information matrix

Following on the simplistic two step procedure of the DCC model, I estimate the parameters in the conditional mean and the conditional volatility specifications by maximizing the sum of log-likelihoods over all the annual series (Jan,...,Dec). Each annual series comprises of the excess returns and the forward rates at yearly interval (e.g. Returns from January each year to January next year along with the forward rates rates in January each year form the $1^{st}$ series). The covariance matrix of the parameters is the inverse of the information matrix times the reciprocal of the sample size. The information matrix is the expected value of the outer products of the scores, usually estimated as the sample mean of the outer product of scores. I take into account the correlation of the scores due to the correlation in the unexpected returns. I use Newey-West type correction to account for the correlation. I use the covariance matrix based on this modified information matrix times the reciprocal of the sample size for inference through out the article.

For example consider the specification pair (S), (S-1).

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)} \;\; \forall\, n = 2, \ldots, 5 \tag{S}$$

$$\epsilon_{t+1}^{(n)} = \sqrt{h_{t+1}^{(n)}} \cdot u_{t+1}^{(n)} \;\; with \;\; u_{t+1}^{(n)} \sim N(0,1)$$

$$h_{t+1}^{(n)} = \exp(\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)}) \tag{S-1}$$

Let, $\Omega$ represent the estimates of the parameters in the conditional mean and the conditional volatility specifications.

$$\Omega = [\kappa^{(n)}, \gamma_{1,\ldots,5}^{(n)}, \alpha^{(n)}, \delta^{(n)}, \beta_{1,\ldots,5}^{(n)}]^{\mathrm{T}}$$

The log-likelihood $\bar{L}(\Omega)$ is the same as log-likelihood in the first step of the DCC model $L_1(\Omega)$. Thus the coefficient estimates are the same. However the time indices are different in this case. The time increment is serial at month's interval, i.e. $1/12^{th}$ time interval. The overlapping data results in the serial correlation, $\text{corr}(\epsilon_{t+1}^{(n)}, \epsilon_{t+1-i}^{(n)})$ for $i = 1/12, 2/12, \ldots, 1$.

$$\bar{L}(\Omega) = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln(2\pi) + \ln(h_t^{(n)}) + \frac{(\epsilon_t^{(n)})^2}{h_t^{(n)}}\right).$$

Let $s_{t+1}$ be the score vector for the log-likelihood conditional on time $t$ information.

$$s_{t+1} = \frac{\partial \ln(f(rx_{t+1}^{(n)}|\Im_t))}{\partial\Omega}$$

$$\ln(f(rx_{t+1}^{(n)}|\Im_t)) = -\frac{1}{2}\left(\ln(2\pi) + \ln(h_{t+1}^{(n)}) + \frac{(\epsilon_{t+1}^{(n)})^2}{h_{t+1}^{(n)}}\right)$$

The score vector $s_{t+1}$ for the specification pair (S), (S-1) is,

$$\frac{\partial h_{t+1}^{(n)}}{\partial\Omega} = h_{t+1}^{(n)}\left[-\frac{2\epsilon_t^{(n)}\delta}{(\epsilon_t^{(n)})^2+\lambda}\left[1, f_{t-1}^{(1,\ldots,5)}\right]_{1\times6}, \left[1, \ln((\epsilon_{t+1}^{(n)})^2+\lambda), f_t^{(1,\ldots,5)}\right]_{1\times7}\right]'_{13\times1}$$

$$\frac{\partial \epsilon_{t+1}^{(n)}}{\partial\Omega} = -\left[\left[1, f_t^{(1,\ldots,5)}\right]_{1\times6}, \mathbf{0}_{1\times7}\right]'_{13\times1}$$

$$s_{t+1} = -\frac{1}{2}\left(\frac{1}{h_{t+1}^{(n)}} - \frac{(\epsilon_{t+1}^{(n)})^2}{(h_{t+1}^{(n)})^2}\right)\frac{\partial h_{t+1}^{(n)}}{\partial\Omega} - \frac{\epsilon_{t+1}^{(n)}}{h_{t+1}^{(n)}}\frac{\partial \epsilon_{t+1}^{(n)}}{\partial\Omega}$$

The multiples of $\frac{\partial h_{t+1}^{(n)}}{\partial\Omega}$, and $\frac{\partial \epsilon_{t+1}^{(n)}}{\partial\Omega}$ contain the unexpected returns $\epsilon_{t+1}^{(n)}$. Also $\frac{\partial h_{t+1}^{(n)}}{\partial\Omega}$ contains functions of unexpected returns $\epsilon_t^{(n)}$ and $\epsilon_{t+1}^{(n)}$. Due to the overlap the unexpected returns are correlated at least up to 12 lags (with $1/12^{th}$ increments in time). Thus the score vectors and the outer products of the scores are likely correlated at least up to 12 lags. The information matrix $\mathcal{I}$ is the expected value of the outer products of the scores. I use Newey-West estimator to obtain the estimate of the information matrix. I use the inverse of this information matrix times the reciprocal of the sample size as the estimate of the covariance matrix of the coefficients. I obtain the information matrix with $q = 12$ and $q = 18$ lags. I do not find any noticeable differences in the standard errors and inference with the increased lags.

$$\hat{\mathcal{I}} \approx \psi_0 + \sum_{\nu=1,\,2,\ldots}^{q} \left(1 - \frac{\nu}{q+1}\right)(\psi_\nu + \psi_{-\nu})$$

$$\text{Where } \psi_\nu = \frac{1}{T} \sum_{t=1}^{T-\frac{\nu}{12}} s_t \cdot s'_{t+\frac{\nu}{12}}$$

$$\widehat{\text{cov}}(\Omega - \hat{\Omega}) \approx T^{-1}\hat{\mathcal{I}}^{-1}$$

The sample autocorrelation function for the unexpected returns $\epsilon_{t+1}^{(n)}$ for bonds of 2-year maturity is in Table 15.

Sample ACF for $\epsilon_{t+1}^{(n)}$ up to 12 lags, lags are at 1/12 interval

| | 1/12 | 2/12 | 3/12 | 4/12 | 5/12 | 6/12 | 7/12 | 8/12 | 9/12 | 10/12 | 11/12 | 12/12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 2$ | 0.87 | 0.78 | 0.70 | 0.62 | 0.56 | 0.50 | 0.44 | 0.40 | 0.35 | 0.28 | 0.22 | 0.14 |
| $n = 3$ | 0.85 | 0.75 | 0.66 | 0.58 | 0.51 | 0.45 | 0.38 | 0.34 | 0.30 | 0.22 | 0.17 | 0.08 |
| $n = 4$ | 0.85 | 0.74 | 0.64 | 0.55 | 0.47 | 0.41 | 0.34 | 0.30 | 0.25 | 0.18 | 0.12 | 0.03 |
| $n = 5$ | 0.83 | 0.71 | 0.62 | 0.52 | 0.44 | 0.37 | 0.29 | 0.26 | 0.21 | 0.13 | 0.08 | -0.01 |

Table 15: The sample autocorrelation function of the unexpected excess returns $\epsilon_{t+1}^{(n)}$ up to 12 lags for bonds of 2- to 5-year maturities. The unexpected returns correspond to the specification pair: (S),(S-1).

# Appendix B

In this section I reduce the number of forward rate spreads used for predicting the conditional volatility. I only keep those spreads that contribute the most toward the conditional volatility estimation. Specification (S-2) contains the short rate and all the four spreads.

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)} \;\; \forall\, n = 2, \ldots, 5 \tag{S}$$

$$h_{t+1}^{(n)} = \exp(\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} \cdot f_t^{(1)} + \beta_{2,\ldots,5}^{(n)} \cdot (f_t^{(2,\ldots,5)} - f_t^{(1)})) \tag{S-2}$$

I refer to the spread between the forward rate with maturity $n$ and the short rate ( $n = 1$ ) as the n-1 spread. Dropping the 5-1 spread results in a significant drop in the log-likelihood values, whereas dropping the 3-1 spread results in no noticeable differences. In fact the coefficients of the 3-1 spread are the least significant of all the spread coefficients in the full level-and-spread specification (S-2) (Table 22). Therefore additional experimentation is left to 4-1 and 2-1 spreads.

$$h_{t+1}^{(n)} = \exp\left[\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} \cdot f_t^{(1)} + \beta_{2,4,5}^{(n)} \cdot (f_t^{(2,4,5)} - f_t^{(1)})\right] \tag{S-5}$$

$$h_{t+1}^{(n)} = \exp\left[\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} \cdot f_t^{(1)} + \beta_{4,5}^{(n)} \cdot (f_t^{(4,5)} - f_t^{(1)})\right] \tag{S-6}$$

$$h_{t+1}^{(n)} = \exp\left[\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} \cdot f_t^{(1)} + \beta_{2,5}^{(n)} \cdot (f_t^{(2,5)} - f_t^{(1)})\right] \tag{S-7}$$

I consider three specifications (S-5), (S-6) and (S-7). (S-5) contains the three spreads 2-1, 4-1 and 5-1, whereas (S-6) contains the 4-1 and 5-1 spreads and the specification (S-7) contains 2-1 and 5-1 spreads.

Table 16 includes the log likelihood values for the three specifications (S-5), (S-6) and (S-7). Although dropping the 2-1 spread does little harm to the log likelihoods ((S-5) vs. (S-6)), dropping the 4-1 spread significantly reduces the log likelihood values for $n = 4, 5$ ((S-5) vs. (S-7)). The likelihood ratio tests reject the specification (S-7) (2-1 and 5-1 spreads and the short rate) for maturities $n = 4, 5$ in favor of specification (S-5) which contains all

Log likelihood values for different specifications.

| | **S-5** | **S-6** | **S-7** |
|---|---|---|---|
| $n = 2$ | 1335.42 | 1335.27 | 1335.12 |
| $n = 3$ | 1055.71 | 1055.12 | 1055.38 |
| $n = 4$ | 911.85 | 911.11 | 907.00 |
| $n = 5$ | 809.63 | 808.59 | 799.20 |

Table 16: The maximum log likelihood values for different specifications. Conditional mean specification (S). Spreads in different conditional volatility specification are: (S-5): 2-1, 4-1, 5-1, (S-6): 4-1, 5-1, (S-7): 2-1, 5-1.

three spreads (2-1, 4-1 and 5-1) and the short rate.

Joint significance statistic for spreads

| | $\chi^2(\beta_{4,5}^{(n)} = 0)$, **(S-6)** | $\chi^2(\beta_{2,5}^{(n)} = 0)$, **(S-7)** | **1% Crit-**$\chi^2$ |
|---|---|---|---|
| $n = 2$ | 2.13 | 1.36 | 9.21 |
| $n = 3$ | 5.15 | 3.01 | 9.21 |
| $n = 4$ | 8.32 | 6.58 | 9.21 |
| $n = 5$ | 11.15 | 8.94 | 9.21 |

Table 17: The joint significance test of the coefficients of spreads in the conditional volatility specification (S-6) (Conditional mean specification: (S)).

Table 17 tabulates the $\chi^2$ statistics for the joint significance of the coefficients of the spreads. The $\chi^2$ statistic are higher for the specification (S-6) that uses the 4-1 and 5-1 spreads than the $\chi^2$ statistic for the 2-1 and 5-1 spreads in specification (S-7). Thus the $\chi^2$ statistic favors the conditional volatility specification (S-6), which contains the 4-1 and 5-1 spreads and the short rate.

$\sigma((h_{t+1}^{(n)})^{1/2})$

| | **(S-5)** | **(S-6)** | **(S-7)** |
|---|---|---|---|
| $n = 2$ | 0.39 | 0.39 | 0.39 |
| $n = 3$ | 0.72 | 0.72 | 0.73 |
| $n = 4$ | 0.99 | 0.99 | 0.96 |
| $n = 5$ | 1.33 | 1.33 | 1.29 |

Table 18: The standard deviation of the conditional standard deviation implied by different specifications. Conditional mean specification (S). Conditional mean specification (S). Spreads in different conditional volatility specification are: (S-5): 2-1, 4-1, 5-1, (S-6): 4-1, 5-1, (S-7): 2-1, 5-1.

The standard deviation of the conditional standard deviation, $\sigma((h_{t+1}^{(n)})^{1/2})$, do not help differentiate between the specifications. Table 18 tabulates the values for different specifications. However the $\sigma((h_{t+1}^{(n)})^{1/2})$ values are lower for the 2-1 and 5-1 spread combination specification (S-7) for maturities $n = 4$ and 5.

Thus the specification (S-7) that uses 2-1 and 5-1 spreads and the short rate is statistically distinguishable from the specification (S-5) that uses the 2-1, 4-1 and 5-1 spreads. The conditional volatility specification (S-7) is inferior to the specification (S-5). Thus I discard the specification (S-7). I keep the specification (S-6) that uses the 4-1, 5-1 spreads and the short rate to forecast the conditional volaility. The specification (S-6) is statistically indistinguishable from the specification (S-5) and specification (S-2) and at the same time uses only two spreads 4-1 and 5-1.

Thus the reduced conditional volatility specification in terms of the level and spreads is (S-6).

$$h_{t+1}^{(n)} = \exp\left[\alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} \cdot f_t^{(1)} + \beta_{4,5}^{(n)} \cdot (f_t^{(4,5)} - f_t^{(1)})\right] \tag{S-6}$$

# Appendix C

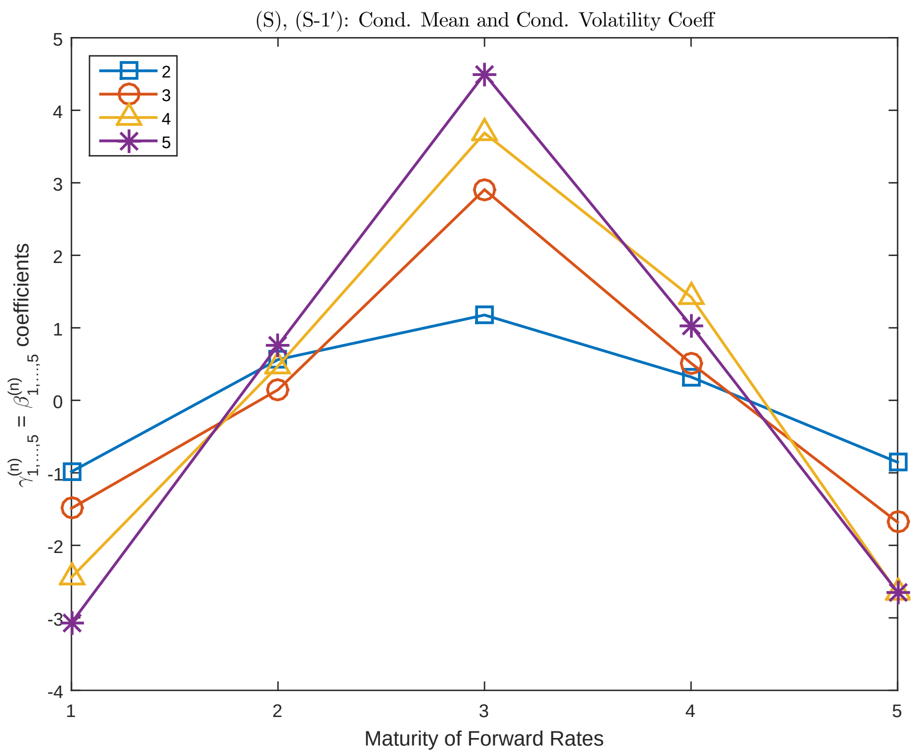


Figure 8: Coefficients of the forward rates in the conditional volatility specification (S-1′) and the conditional mean specification (S). Coefficients of the forward rates in the conditional volatility specification (S-1′) are restricted to be the same as the coefficients in the conditional mean specification (S). Volatility is in % standard deviation.

$$rx^{(n)}_{t+1} = \kappa^{(n)} + \gamma^{(n)}_{1,...,5} \cdot f^{(1,...,5)}_t + \epsilon^{(n)}_{t+1}$$

| | $\kappa^{(n)}$ | $\gamma^{(n)}_1$ | $\gamma^{(n)}_2$ | $\gamma^{(n)}_3$ | $\gamma^{(n)}_4$ | $\gamma^{(n)}_5$ |
|---|---|---|---|---|---|---|
| $n = 2$ | -0.02 | -0.98 | 0.56 | 1.18 | 0.32 | -0.85 |
| | (-6.4) | (-10.2) | (2.6) | (5.7) | (2.7) | (-9.3) |
| $n = 3$ | -0.03 | -1.49 | 0.15 | 2.91 | 0.51 | -1.68 |
| | (-5.9) | (-16.1) | (0.9) | (15.1) | (4.4) | (-12.2) |
| $n = 4$ | -0.04 | -2.44 | 0.47 | 3.69 | 1.43 | -2.66 |
| | (-7.4) | (-9.8) | (0.9) | (13.5) | (9.7) | (-7.5) |
| $n = 5$ | -0.05 | -3.07 | 0.76 | 4.50 | 1.03 | -2.64 |
| | (-7.6) | (-25.5) | (18.4) | (12.2) | (6.1) | (-15.3) |

Table 19: The Specification pair: (S),(S-1′). The coefficient estimates and the t-ratios in brackets for the coefficients of the forward rates in the conditional mean specification (S). The coefficients of the forward rates in the conditional mean and the conditional volatility specification are the same for this specification pair.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)}$$

| | $\alpha^{(n)}$ | $\delta_1^{(n)}$ | $\gamma_1^{(n)}$ | $\gamma_2^{(n)}$ | $\gamma_3^{(n)}$ | $\gamma_4^{(n)}$ | $\gamma_5^{(n)}$ |
|---|---|---|---|---|---|---|---|
| $n = 2$ | -7.95 | 0.04 | -0.98 | 0.56 | 1.18 | 0.32 | -0.85 |
| | (-20.5) | (1.0) | (-10.2) | (2.6) | (5.7) | (2.7) | (-9.3) |
| $n = 3$ | -6.55 | 0.07 | -1.49 | 0.15 | 2.91 | 0.51 | -1.68 |
| | (-20.1) | (1.8) | (-16.1) | (0.9) | (15.1) | (4.4) | (-12.2) |
| $n = 4$ | -6.10 | 0.06 | -2.44 | 0.47 | 3.69 | 1.43 | -2.66 |
| | (-22.2) | (1.8) | (-9.8) | (0.9) | (13.5) | (9.7) | (-7.5) |
| $n = 5$ | -5.61 | 0.07 | -3.07 | 0.76 | 4.50 | 1.03 | -2.64 |
| | (-23.5) | (2.9) | (-25.5) | (18.4) | (12.2) | (6.1) | (-15.3) |

Table 20: The Specification pair: (S),(S-1′). The coefficient estimates and the t-ratios in brackets for the coefficients of the forward rates in the conditional mean specification (S-1′). The coefficients of the forward rates in the conditional mean and the conditional volatility specification are the same for this specification pair.

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)}$$

| | $\kappa^{(n)}$ | $\gamma_1^{(n)}$ | $\gamma_2^{(n)}$ | $\gamma_3^{(n)}$ | $\gamma_4^{(n)}$ | $\gamma_5^{(n)}$ |
|---|---|---|---|---|---|---|
| n = 2 | -0.02 | -0.83 | 0.15 | 1.19 | 0.46 | -0.76 |
| | (-5.0) | (-6.1) | (0.6) | (5.0) | (2.4) | (-5.0) |
| $n = 3$ | -0.03 | -1.72 | -0.02 | 3.09 | 0.59 | -1.64 |
| | (-4.9) | (-7.7) | (-0.0) | (7.7) | (1.8) | (-6.8) |
| $n = 4$ | -0.04 | -2.20 | -0.15 | 3.59 | 1.54 | -2.35 |
| | (-5.3) | (-6.9) | (-0.2) | (6.1) | (3.2) | (-6.5) |
| $n = 5$ | -0.05 | -2.88 | 0.26 | 3.97 | 1.53 | -2.35 |
| | (-5.5) | (-7.5) | (0.4) | (5.5) | (2.6) | (-5.4) |

Table 21: The Specification pair: (S),(S-2). The coefficient estimates and the t-ratios in brackets for the coefficients of the forward rates in the conditional mean specification (S).

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{2,\ldots,5}^{(n)} \cdot (f_t^{(2,\ldots,5)} - f_t^{(1)})$$

| | $\alpha^{(n)}$ | $\delta_1^{(n)}$ | $\beta_1^{(n)}$ | $\beta_2^{(n)}$ | $\beta_3^{(n)}$ | $\beta_4^{(n)}$ | $\beta_5^{(n)}$ |
|---|---|---|---|---|---|---|---|
| n = 2 | -10.03 | -0.04 | 18.00 | 11.71 | 4.75 | -13.36 | 9.13 |
| | (-17.8) | (-1.1) | (4.5) | (0.5) | (0.2) | (-0.7) | (0.7) |
| $n = 3$ | -9.17 | -0.06 | 19.28 | 23.99 | 5.15 | -12.80 | 12.32 |
| | (-20.2) | (-2.0) | (5.0) | (0.9) | (0.2) | (-0.7) | (0.9) |
| $n = 4$ | -8.19 | -0.02 | 18.82 | 28.77 | 3.28 | -10.46 | 12.77 |
| | (-16.7) | (-0.5) | (5.0) | (1.1) | (0.1) | (-0.6) | (0.9) |
| $n = 5$ | -7.89 | -0.02 | 20.18 | 30.25 | 7.07 | -11.06 | 11.83 |
| | (-16.8) | (-0.5) | (5.6) | (1.1) | (0.3) | (-0.6) | (0.9) |

Table 22: The Specification pair: (S),(S-2). The coefficient estimates and the t-ratios in brackets for the coefficients of the forward rates in the conditional volatility specification (S-2).

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)}$$

| | $\kappa^{(n)}$ | $\gamma_1^{(n)}$ | $\gamma_2^{(n)}$ | $\gamma_3^{(n)}$ | $\gamma_4^{(n)}$ | $\gamma_5^{(n)}$ |
|---|---|---|---|---|---|---|
| $n = 2$ | -0.02 | -0.84 | 0.20 | 1.16 | 0.46 | -0.76 |
| | (-5.0) | (-6.4) | (0.8) | (5.1) | (2.4) | (-5.1) |
| $n = 3$ | -0.03 | -1.77 | 0.11 | 3.03 | 0.56 | -1.62 |
| | (-4.9) | (-7.9) | (0.2) | (7.7) | (1.7) | (-6.5) |
| $n = 4$ | -0.04 | -2.30 | 0.15 | 3.44 | 1.49 | -2.35 |
| | (-5.2) | (-7.2) | (0.2) | (6.3) | (3.2) | (-6.4) |
| $n = 5$ | -0.05 | -2.94 | 0.47 | 3.83 | 1.47 | -2.31 |
| | (-5.4) | (-7.6) | (0.6) | (5.6) | (2.5) | (-5.1) |

Table 23: The specification pair: (S), (S-6). The coefficient estimates and the t-ratios in brackets for the conditional mean specification (S).

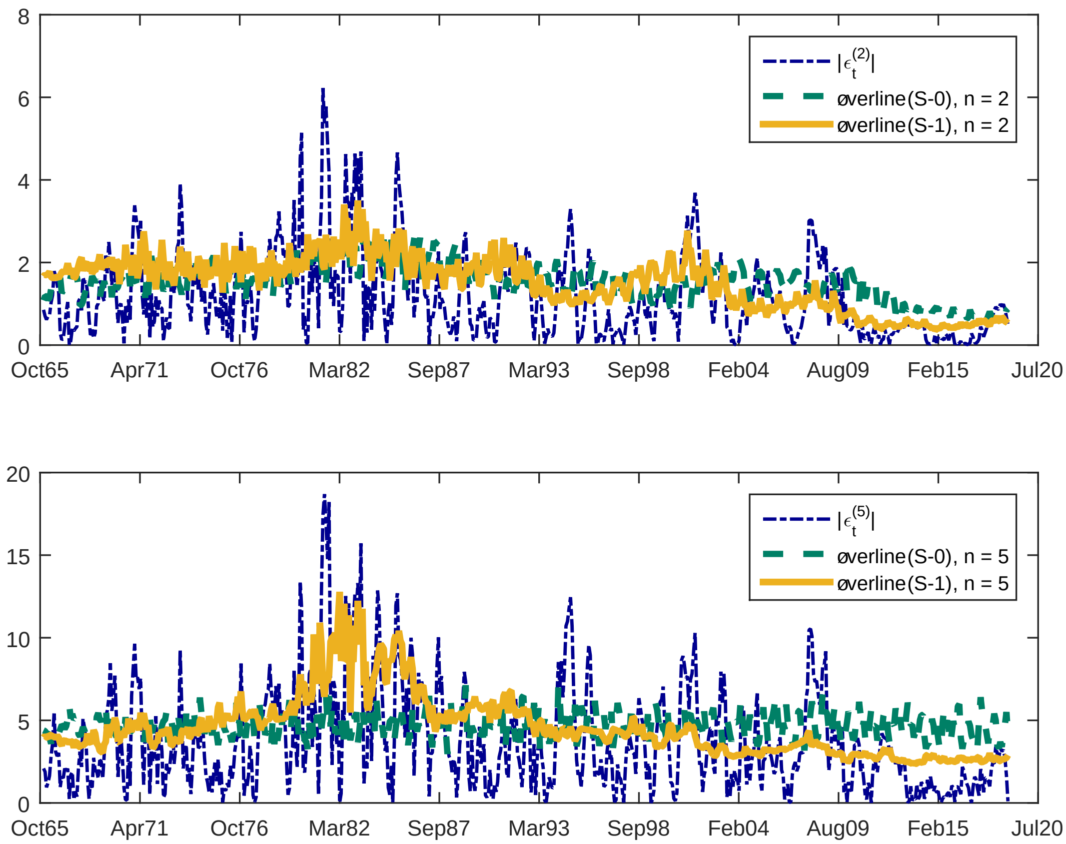


Figure 9: Comparison: Magnitude of unexpected returns , $|\epsilon_t^{(\cdot)}|$, conditional volatility implied by $(\overline{S-0})$ and $(\overline{S-1})$ spec., for excess returns of bonds of maturities $n = 2, 5$. All values in % and conditional volatilities in standard deviation.

# Appendix D

I study term structure drivers of conditional volatility of one year bond returns in the data sample spanning the period from January, 1964 to December, 1979. This analysis is an exercise to verify whether the conclusions drawn using full sample of data hold true in pre-80s data.

I find that the conclusions drawn with the full data sample are still valid in the limited dataset. The reason for this congruence is the fact that the interest rate level is the main driver of conditional volatility of one year bond returns. Given a single dominant driver, it can be estimated with precision even in a shorter data sample and hence the congruence between the results with shorter and full data sample.

I also forecast conditional volatility following the typical methods used by forecasters. This implies using only the data known to a forecaster at the time of forecast for formulating the forecast. I create two separate series of forecasts. One uses the expanding window method and the other rolling window method. I present comparisons of these forecasts with the full sample estimate of conditional volatility. Both of these methods yield a seemingly close match with the full sample value of conditional volatility. This match is owed to the dominance of interest rate level in capturing the variation in conditional volatility. These results imply minimal impact of forward looking bias on full sample estimates.

## D.1 Econometric analysis with data limited to December, 1979

To ease the reading of analysis presented below, first I list the complete formulas of specifications from the main article. I compute maximum likelihood estimates of the coefficients in specification pair (S), (S-1) using the limited data sample spanning the period January, 1964 to December, 1979.

$$rx_{t+1}^{(n)} = \kappa^{(n)} + \gamma_{1,\dots,5}^{(n)} \cdot f_t^{(1,\dots,5)} + \epsilon_{t+1}^{(n)} \;\; \forall\, n = 2, \dots, 5 \tag{S}$$

$$\epsilon_{t+1}^{(n)} = \sqrt{h_{t+1}^{(n)}} \cdot u_{t+1}^{(n)} \;\; with \;\; u_{t+1}^{(n)} \sim N(0,1)$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) \tag{S-0}$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_{1,\dots,5}^{(n)} \cdot f_t^{(1,\dots,5)} \tag{S-1}$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \gamma_{1,\dots,5}^{(n)} \cdot f_t^{(1,\dots,5)} \tag{S-1$'$}$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{2,\dots,5}^{(n)} \cdot (f_t^{(2,\dots,5)} - f_t^{(1)}) \tag{S-2}$$

Table 24 presents the coefficient estimates from conditional volatility specification (S-1) and Table 13 (in main article) includes statistical tests on these coefficients. Similar to the full-sample case, the coefficients of forward rates in (S-1) are not individually significant (Table 24), but jointly significant (first row in Table 13). The third row in Table 13 includes likelihood ratio test comparing (S-0) and (S-1). The specification (S-0) which lacks any term structure information is summarily rejected in favor of (S-1). This result reveals that, similar to the full sample case, the term structure captures a non-trivial component of the variation in conditional volatility. The last row in Table 13 also presents comparison between (S-1′) and (S-1). (S-1′) which sets same coefficients for the forward rates in conditional volatility specification as that in conditional mean ($\gamma_{1,\dots,5}^{(n)} = \beta_{1,\dots,5}^{(n)}$) is completely rejected in favor of specification (S-1). This test, similar to the full sample case, states that the term structure drivers of the conditional mean and the conditional volatility are not the same.

Note that the hypothesis $\sum_{i=1}^{5} \beta_i^{(n)} = 0$ for $\beta_i^{(n)}$ in (S-1), which tests for statistical *insignificance* of the level of interest rates, is resoundingly rejected (second row in Table 13). Therefore, similar to the full sample case, I split term structure drivers of conditional volatility into level and spread components and analyze their importance. For this purpose I estimate the coefficients in (S-2) which separates interest rate level variables from those for the spreads. Table 25 lists these estimates.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_{1,\dots,5}^{(n)} \cdot f_t^{(1,\dots,5)}$$

| | $\alpha^{(n)}$ | $\delta_1^{(n)}$ | $\beta_1^{(n)}$ | $\beta_2^{(n)}$ | $\beta_3^{(n)}$ | $\beta_4^{(n)}$ | $\beta_5^{(n)}$ |
|---|---|---|---|---|---|---|---|
| $n = 2$ | -12.87 | -0.14 | -9.12 | 32.45 | 8.85 | -19.93 | 19.91 |
| | (-7.7) | (-1.8) | (-1.3) | (0.9) | (0.9) | (-0.7) | (1.2) |
| $n = 3$ | -10.81 | -0.09 | -15.11 | 37.20 | 9.65 | -17.80 | 18.53 |
| | (-7.6) | (-0.9) | (-0.9) | (0.9) | (0.3) | (-0.6) | (0.7) |
| $n = 4$ | -8.94 | -0.01 | -40.50 | 48.27 | 9.16 | -14.90 | 22.59 |
| | (-7.4) | (-0.0) | (-0.7) | (0.6) | (0.5) | (-0.5) | (0.7) |
| $n = 5$ | -8.32 | 0.01 | -36.93 | 43.22 | 13.88 | -12.01 | 23.04 |
| | (-7.7) | (0.4) | (-0.6) | (0.7) | (0.6) | (-0.4) | (0.5) |

Table 24: The coefficient estimates and the t-ratios in brackets for the conditional volatility part of the specification (S-1) from the specification pair (S), (S-1) using data until December, 1979.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{2,\dots,5}^{(n)} \cdot (f_t^{(2,\dots,5)} - f_t^{(1)})$$

| | $\alpha^{(n)}$ | $\delta_1^{(n)}$ | $\beta_1^{(n)}$ | $\beta_2^{(n)}$ | $\beta_3^{(n)}$ | $\beta_4^{(n)}$ | $\beta_5^{(n)}$ |
|---|---|---|---|---|---|---|---|
| $n = 2$ | -12.87 | -0.14 | 32.15 | 32.45 | 8.85 | -19.93 | 19.91 |
| | (-7.7) | (-1.8) | (4.2) | (0.9) | (0.9) | (-0.7) | (1.2) |
| $n = 3$ | -10.81 | -0.09 | 32.46 | 37.20 | 9.65 | -17.80 | 18.53 |
| | (-7.6) | (-0.9) | (3.9) | (0.9) | (0.3) | (-0.6) | (0.7) |
| $n = 4$ | -8.94 | -0.01 | 24.62 | 48.27 | 9.16 | -14.90 | 22.59 |
| | (-7.4) | (-0.0) | (2.9) | (0.6) | (0.5) | (-0.5) | (0.7) |
| $n = 5$ | -8.32 | 0.01 | 31.20 | 43.22 | 13.88 | -12.01 | 23.04 |
| | (-7.7) | (0.4) | (2.7) | (0.7) | (0.6) | (-0.4) | (0.5) |

Table 25: The coefficient estimates and the t-ratios in brackets for the conditional volatility part of the specification (S-2) from the specification pair (S), (S-2) using data until December, 1979.

The full sample analysis shows that $\beta_1^{(n)}$ in (S-2) is strongly significant and as a result level is the most important driver of conditional volatility. Similarly, Table 25 shows that this is true even in the limited sample of data covering the period January, 1964 to December, 1979. $\beta_1^{(n)}$ is significant even in this short data sample. The relatively smaller (but still greater than 2.5) t-ratios for $\beta_1^{(n)}$ in Table 25 are likely to be due to smaller sample size.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{2,\ldots,5}^{(n)} \cdot (f_t^{(2,\ldots,5)} - f_t^{(1)}) \tag{S-2}$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_{2,\ldots,5}^{(n)} \cdot (f_t^{(2,\ldots,5)} - f_t^{(1)}) \tag{S-3}$$

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} \tag{S-4}$$

Conditional volatility specifications: S-2, S-3 and S-4

| $\chi^2 : H_1$ | $n = 5$ | $n = 4$ | $n = 3$ | $n = 2$ |
|---|---|---|---|---|
| $\chi^2$ **(S-3 vs S-2): $H_1$: S-2** | 14.28 | 14.71 | 23.25 | 27.03 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| $\chi^2$ **(S-4 vs S-2): $H_1$: S-2** | 7.54 | 8.04 | 15.27 | 23.11 |
| | (0.11) | (0.09) | (0.00) | (0.00) |

Table 26: The table presents $\chi^2$ statistics and P-values in brackets for likelihood ratio tests comparing conditional volatility specifications (S-3 vs S-2) and (S-4 vs S-2). The same conditional mean specification (S) applies for both the tests. The estimates are based on data spanning till December, 1979.

The importance of interest rate level is also reflected in the likelihood ratio test comparing specification (S-3) with (S-2). The specification (S-3) is in fact specification (S-2) but minus the variable for interest rate level. The likelihood ratio test in the first row in Table 26 strongly rejects (S-3) in favor of (S-2). This result highlights the importance of interest rate level for capturing variation in the conditional volatility, even in the limited data sample extending till December, 1979.

The second row in Table 26 tests for importance of spreads by comparing specification (S-4) with (S-2). The specification (S-4) is nested within (S-2). Specification (S-4) is (S-2) but with the coefficients of spreads constrained to zero. The test results in Table 26 present mixed support for spreads in (S-2). The spreads do not appear to be that important for capturing variation in conditional volatility for $n = 4, 5$, but quite important for $n = 2, 3$. The full sample tests yield similar results but with opposing patterns. In the case of full sample, the spreads are essential part of specification (S-2) for $n = 4, 5$ but not an essential part for the case of $n = 2, 3$.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{4,5}^{(n)} \cdot (f_t^{(4,5)} - f_t^{(1)}) \qquad \text{(S-6)}$$

Similar to the exercise with full sample, I reduce the number of spreads in specification (S-2) to two. The 5-1 spread is the most important and removing related term lowers the log-likelihood to a limited extent. Of the remaining spreads, 4-1, 3-1 and 2-1, the statistical tests do not provide strong guidance for picking one over the other. However, in the full sample case the tests decidedly point towards including 4-1 spread in the specification. Consequently, I keep 4-1 spread in addition to 5-1 spread in the reduced specification (S-6) to facilitate comparison with full sample case. The estimates of the coefficients in this reduced form specification (S-6) are presented in Table 27.

$$\ln(h_{t+1}^{(n)}) = \alpha^{(n)} + \delta_1^{(n)} \cdot \ln((\epsilon_t^{(n)})^2 + \lambda) + \beta_1^{(n)} f_t^{(1)} + \beta_{4,5}^{(n)} \cdot (f_t^{(4,5)} - f_t^{(1)})$$

| | $\alpha^{(n)}$ | $\delta_1^{(n)}$ | $\beta_1^{(n)}$ | $\beta_4^{(n)}$ | $\beta_5^{(n)}$ |
|---|---|---|---|---|---|
| $n = 2$ | -12.68<br>(-7.9) | -0.13<br>(-1.8) | 29.97<br>(4.4) | -12.27<br>(-0.8) | 13.27<br>(1.4) |
| $n = 3$ | -11.22<br>(-8.6) | -0.12<br>(-0.8) | 18.81<br>(4.3) | -7.80<br>(-0.8) | 14.73<br>(1.3) |
| $n = 4$ | -9.43<br>(-9.4) | -0.03<br>(-0.3) | 27.93<br>(3.5) | -6.43<br>(-0.6) | 15.61<br>(1.3) |
| $n = 5$ | -8.88<br>(-9.8) | -0.03<br>(-0.3) | 29.64<br>(3.1) | -4.79<br>(-0.6) | 17.20<br>(1.5) |

Table 27: The coefficient estimates and the t-ratios in brackets for the conditional volatility part of the specification (S-6) from the specification pair (S), (S-6) using data until December, 1979.

The above analysis shows that (1) the term structure of interest rates captures substantial amount of variation in the conditional volatility of one year bond returns; (2) term structure drivers of the conditional mean and the conditional volatility are different; (3) the overall level of interest rates are essential for capturing variation in the conditional volatilities; (4) the spreads however do not have a clear or strong support for their importance in explaining the conditional volatilities. Further, in light of above determinations Cochrane-Piazzesi factor captures aggregate price of risk and not the amount of bond return risk even in the data sample limited to December, 1979. Thus the conclusions arrived at using the full sample of data remain unaltered when the analysis is conducted with limited sample of data extending till December, 1979.

## D.2 Graphs of coefficient estimates from conditional volatility specification

I graph the coefficient estimates of forward rates and/or related spreads from specifications (S-1), (S-2) and (S-6) to verify whether the estimates of each of these specifications share a common pattern/component across different maturities of bonds ($n = 2, 3, 4, 5$) and whether this pattern is similar to that obsereved with the full sample analysis. To this end, I graph these coefficient estimates in Fig. 10. The figure depicts a common sideway-Z pattern/component for the coefficients in (S-1) and (S-2) specifications and a common V-shaped pattern for the coefficients in (S-6). Further, these patterns are similar to those observed with the entire data sample (right-hand panels in Fig. 3 and 7 in the main article).

## D.3 Forecasting conditional volatility

I conduct forecasting exercise for the conditional volatilities. In addition to out-of-sample analysis, forecasting exercise provides a way to gauge the impact of forward looking bias on conditional volatility. The full sample conditional volatility estimate is the baseline in this process and the deviations of forecasts from full sample values measure the impact of

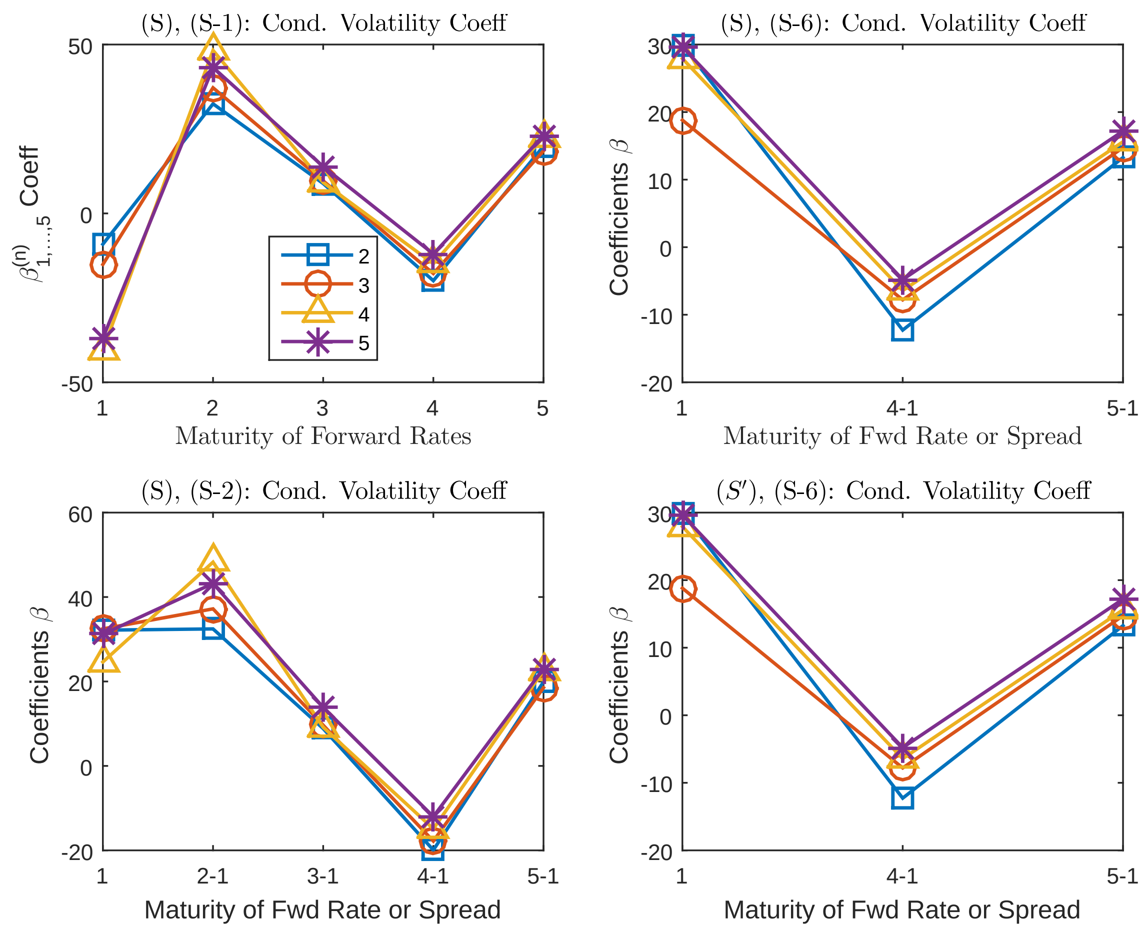


Figure 10: Coefficients of forward rates and spreads from different conditional volatility specifications. (Top-Left): specification (S-1). (Top-Right): specification (S-6). (Bottom-Left): specification (S-2). (Bottom-Right): specification (S-6). The same conditional mean specification (S) applies to all except for the Bottom-Right graph for which ($S'$) is the applicable conditional mean specification. Specification ($S'$): $rx^{(n)}_{t+1} = \kappa^{(n)} + \gamma^{(n)}_1 f^{(1)}_t + \gamma^{(n)}_{2,...,5} \cdot (f^{(2,...,5)}_t - f^{(1)}_t) + \epsilon^{(n)}_{t+1} \;\; \forall \, n = 2, \ldots, 5$

forward looking bias. The referee's suggestion is to mimic forecaster's method. I do this first by using expanding window method and then the rolling window method. I conduct forecasting exercise using specification pair (S) and (S-1).

The expanding window method involves using data available until time t-1 to forecast conditional volatility of returns over t-1 to t. I do so, to reliably estimate coefficients of 13 variables that define the mean (S) and volatility (S-1) specifications. Since coefficients are estimated using the data until December, 1977, the first available forecast of the conditional volatility is for the returns over January, 1978 to December, 1978. Thereafter as the estimation window expands the forecasts are available for each future time period.

I graph the contemporaneous full sample and forecasted values of conditional volatilities in Fig. 11. For clarity the figure includes graphs only for $n = 2, 5$. Similar close associations between full sample estimates and forecasted values apply for $n = 3, 4$ as well. Also, note that as time progresses, expanding window covers more data and the forecasted values of conditional volatility get closer to that computed using full sample of data. The regressions of full sample conditional volatilities on forecasted values yield $R^2$s of about 82%[1].

The $R^2$s seem to be high. The reason, however, for such a close match is the importance of interest rate level in capturing the variation in conditional volatility. Thus although individual coefficient estimates in specification (S-1) can be imprecise, but if the sum of these coefficients ($\sum_{i=1}^{5} \beta_i^{(n)}$) is estimated accurately, it is sufficient for capturing a large portion of the variation in conditional volatility. Similar to the full sample result, Table 13 shows that this sum has strong statistical significance (hence very small standard errors compared to the point estimate) even in the limited data sample. Therefore the impact of interest level is well estimated even with a shorter sample of data. In case specification (S-2) were to be used for conditional volatility, it will be the coefficient $\beta_1^{(n)}$ which would capture the impact of interest rate level. Table 25 shows that this coefficient is strongly significant and as a result it is precisely estimated even in the short sample of data. In

[1] To be specific the values of $R^2$ are 82%, 79%, 84% and 84% for $n = 2, 3, 4$ and 5 respectively.

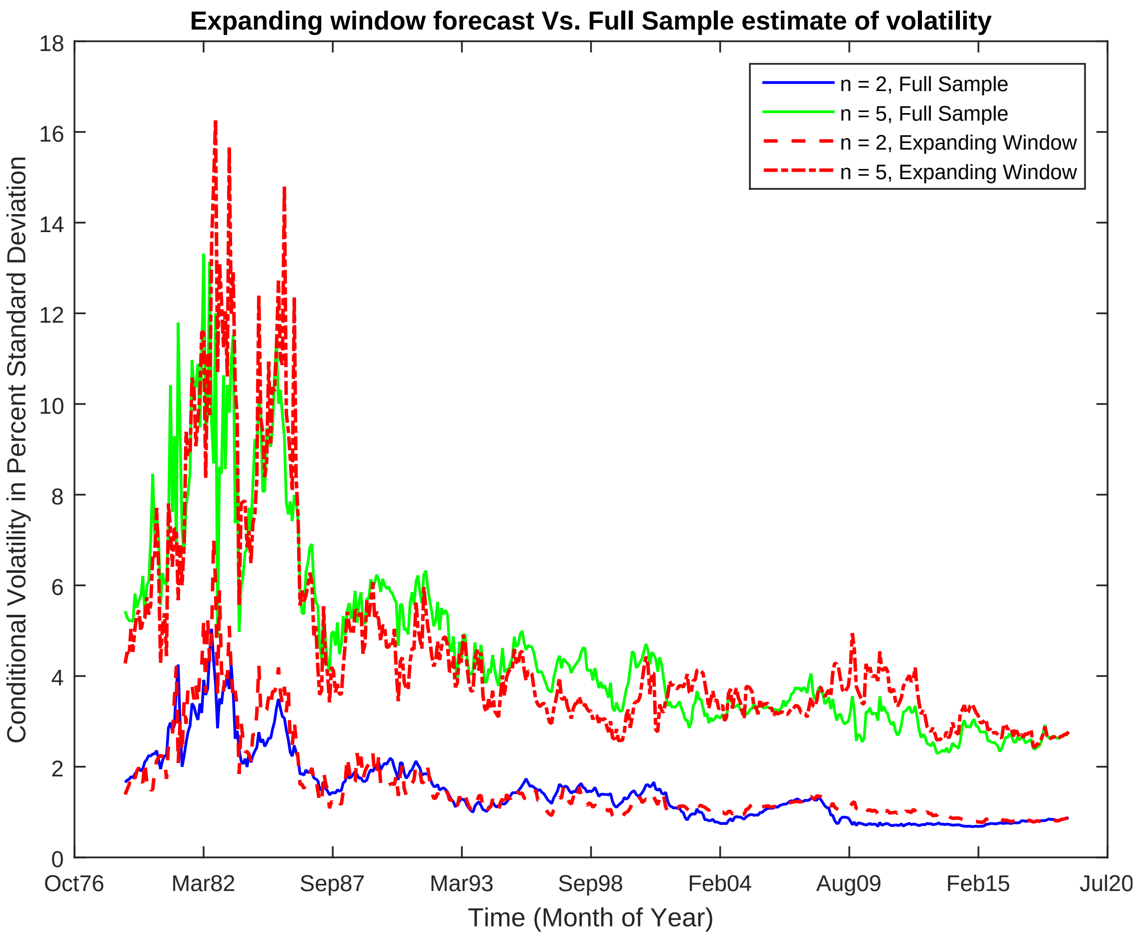


Figure 11: Comparison between the conditional volatility estimated using the full sample of data (solid lines) with the forecasted conditional volatility computed using an expanding window approach (dash and dash-dot lines) using specification pair (S), (S-1). The expanding window forecast utilizes data available until time t-1 to forecast conditional volatility of returns over t-1 to t. For clarity the graph presents these comparisons only for $n = 2, 5$. The conditional volatility values on Y-axis are in %. For first forecast the coefficients are estimated using the data covering the period from January, 1964 to December, 1977. The first forecast for conditional volatility is available for returns over January, 1978 to December, 1978. The data window expands for each future forecast.

addition the coefficient estimates get closer to full sample estimates as more and more data is incorporated for computing the forecast. Consequently, these two aspects combined result in a close match between the foretasted and the full sample values of conditional volatility in Fig. 11. In summary, it is safe to conclude that the impact of forward looking bias on conditional volatility estimates is minimal.

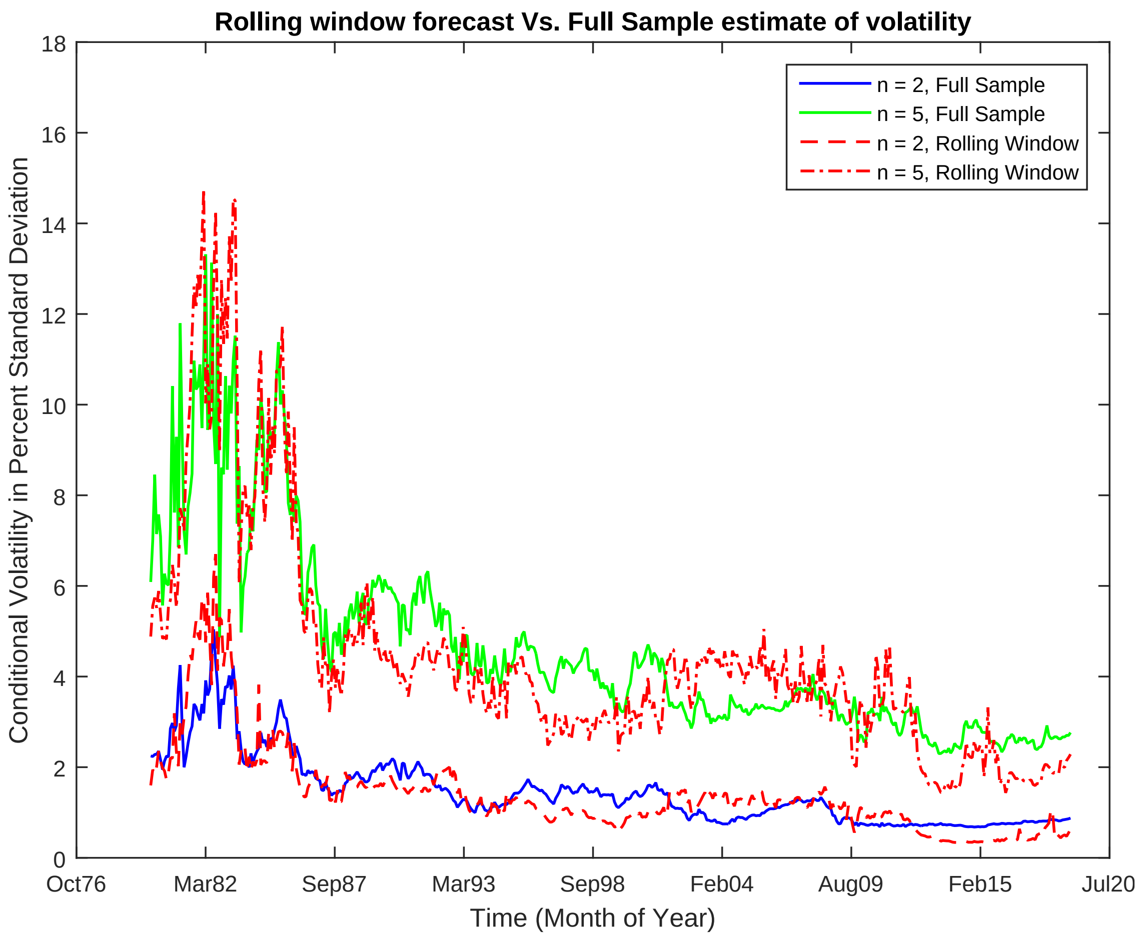


Figure 12: Comparison between the conditional volatility estimated using the full sample of data (solid lines) with the forecasted conditional volatility computed using a rolling window approach (dash and dash-dot lines) using specification pair (S), (S-1). The rolling window forecast utilizes data available from t-15 to t-1 to forecast conditional volatility of returns over t-1 to t. For clarity the graph presents these comparisons only for $n = 2, 5$. The conditional volatility values on Y-axis are in %. For first forecast the coefficients are estimated using the data covering the period from January, 1964 to December, 1978. The first forecast for conditional volatility is available for returns over January, 1979 to December, 1979. The data window rolls with each future forecast.

I also compute conditional volatility forecast using rolling window approach, where I use past 15-years of data to produce the forecast for conditional volatility of next year's return. This method uses the same length of data sample for coefficient estimation. Consequently, the differences between rolling forecast and full sample estimate of conditional volatility can either be due to potential regime changes or estimation errors emerging from use of short data sample in the estimation step. Therefore the comparisons using rolling window approach must be evaluated with care.

Fig. 12 compares the rolling window forecast with full sample estimate of conditional volatility. The $R^2$s from the regression of full sample conditional volatilities on rolling window forecasts are about 78%[2]. These $R^2$s are high considering that only 15-years of data is used at any time to compute the conditional volatility forecast. The seemingly high $R^2$s reaffirm the earlier thesis about the importance of interest rate level for forecasting the conditional volatility. The high $R^2$s provide confidence in the full sample estimate of conditional volatility and related model coefficients.

[2]The individual $R^2$ values are 78%, 77%, 77% and 81% for $n = 2, 3, 4$ and 5 respectively.

# Appendix E

I provide additional analysis with standard ARCH/GARCH models in this appendix. I find that these models also indicate that the lagged conditional volatility does not improve the ability to forecast the conditional volatility. I list the specifications for conditional mean and ARCH/GARCH models of conditional volatility below. Note that the unit time interval is one year. Less than a year's time intervals are represented in fractions of 12.

$$
\begin{aligned}
rx_{t+1}^{(n)} &= \kappa^{(n)} + \gamma_{1,\ldots,5}^{(n)} \cdot f_t^{(1,\ldots,5)} + \epsilon_{t+1}^{(n)} \;\; \forall\, n = 2, \ldots, 5 && \text{(S)} \\
\epsilon_{t+1}^{(n)} &= \sqrt{h_{t+1}^{(n)}} \cdot u_{t+1}^{(n)} \;\; with \;\; u_{t+1}^{(n)} \sim N(0,1) \\
h_{t+1}^{(n)} &= \alpha^{(n)} + \eta_1^{(n)} \cdot h_t^{(n)} + \eta_2^{(n)} \cdot h_{t-1/12}^{(n)} + \delta_1^{(n)} \cdot (\epsilon_t^{(n)})^2 + \delta_2^{(n)} \cdot (\epsilon_{t-1/12}^{(n)})^2, && \text{(B-1)} \\
h_{t+1}^{(n)} &= \alpha^{(n)} + \eta_1^{(n)} \cdot h_t^{(n)} + \delta_1^{(n)} \cdot (\epsilon_t^{(n)})^2, && \text{(B-2)} \\
h_{t+1}^{(n)} &= \alpha^{(n)} + \delta_1^{(n)} \cdot (\epsilon_t^{(n)})^2. && \text{(B-3)}
\end{aligned}
$$

Table 28 presents the estimates of the coefficients for the GARCH(2,2) specification (B-1) and the corresponding t-ratios. The monthly overlap of 1-year excess returns results in correlated unexpected returns. The estimation and the standard error correction procedure is the same as that used for other specifications. The standard errors are based on the modified information matrix that accounts for the correlation of unexpected returns and the scores (Appendix A). The t-ratios indicate that except the square of the first lagged residual $(\epsilon_t^{(n)})^2$, other terms in the specification do not contribute much toward forecasting the conditional variances. In fact the joint significance test does not reject the hypothesis that the coefficients $\eta_1^{(n)}$, $\eta_2^{(n)}$ and $\delta_2^{(n)}$ are zero. Likelihood ratios tests of the simplified specifications, (B-2), a GARCH(1,1) specification, and (B-3) an ARCH(1), specification, indicate that in fact the GARCH(2,2) specification is not distinguishable from the GARCH(1,1) or the ARCH(1) specifications (Table 29).

$$h_{t+1}^{(n)} = \alpha^{(n)} + \eta_1^{(n)} \cdot h_t^{(n)} + \eta_2^{(n)} \cdot h_{t-1/12}^{(n)} + \delta_1^{(n)} \cdot (\epsilon_t^{(n)})^2 + \delta_2^{(n)} \cdot (\epsilon_{t-1/12}^{(n)})^2$$

| | $\alpha^{(n)} \times 100$ | $t(\alpha^{(n)})$ | $\eta_1^{(n)}$ | $t(\eta_1^{(n)})$ | $\eta_2^{(n)}$ | $t(\eta_2^{(n)})$ | $\delta_1^{(n)}$ | $t(\delta_1^{(n)})$ | $\delta_2^{(n)}$ | $t(\delta_2^{(n)})$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $n = 2$ | 0.01 | 1.19 | 0.15 | 0.22 | 0.39 | 0.54 | 0.09 | 1.20 | 0.02 | 0.31 |
| $n = 3$ | 0.04 | 1.46 | 0.10 | 0.17 | 0.33 | 0.56 | 0.11 | 1.54 | 0.03 | 0.47 |
| $n = 4$ | 0.08 | 1.61 | 0.11 | 0.19 | 0.29 | 0.52 | 0.13 | 1.91 | 0.02 | 0.38 |
| $n = 5$ | 0.12 | 1.98 | 0.11 | 0.25 | 0.28 | 0.62 | 0.13 | 1.79 | 0.06 | 1.01 |

Table 28: Estimates of the coefficients for GARCH(2,2) specification (B-1) and the t-ratios.

$H_0$ : B-3 Vs. $H_1$ : B-1

| | $\chi^2$ | $P - Value$ | $Crit - \chi^2$ |
|---|---|---|---|
| $n = 2$ | 2.98 | 0.39 | 7.81 |
| $n = 3$ | 2.59 | 0.46 | 7.81 |
| $n = 4$ | 1.72 | 0.63 | 7.81 |
| $n = 5$ | 3.64 | 0.30 | 7.81 |

Table 29: The $\chi^2$ statistic for the likelihood ratio test of ARCH(1) specification (B-3) against the alternate GARCH(2,2) specification (B-1). Crit-$\chi^2$ is for 5% significance level.

$$h_{t+1}^{(n)} = \alpha^{(n)} + \delta_1^{(n)} \cdot (\epsilon_t^{(n)})^2$$

| | $\alpha^{(n)} \times 100$ | $t(\alpha^{(n)})$ | $\delta_1^{(n)}$ | $t(\delta_1^{(n)})$ |
|---|---|---|---|---|
| $n = 2$ | 0.02 | 13.12 | 0.09 | 1.40 |
| $n = 3$ | 0.07 | 13.00 | 0.13 | 2.08 |
| $n = 4$ | 0.12 | 12.62 | 0.16 | 2.54 |
| $n = 5$ | 0.18 | 12.28 | 0.18 | 2.88 |

Table 30: Coefficients from the ARCH(1) specification (B-3) for the conditional volatility of bond excess returns and the corresponding t-ratios.

The coefficient estimates and the t-ratios for the ARCH(1) specifications are in Table 30. The standard deviation of the conditional standard deviation, $\sigma((h_{t+1}^{(n)})^{1/2})$, values are [0.12, 0.30, 0.45, 0.65] for the GARCH(2,2) model and [0.11, 0.29, 0.43, 0.63] for the ARCH(1) model for bonds of maturities $n = 2, 3, 4, 5$ respectively. The two values are not much different. Thus, even in the context of ARCH/GARCH models, *the lagged conditional variances do not significantly improve the ability to forecast the conditional volatility.* Also, note that the $\sigma((h_{t+1}^{(n)})^{1/2})$ values [0.39, 0.73, 0.99, 1.34] for specification (S-1) are way higher (2 to 3 times) compared to that for the GARCH(2,2) or ARCH(1) specifications. Thus the log conditional volatility specifications with the term structure information capture substantially higher amounts of the variation in the conditional volatilities than the ARCH/GARCH-type models.

# References

Engle, R., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339–350.