Affine Term Structure Models, Volatility and the Segmentation Hypothesis

Kris Jacobs Lotfi Karoui
Desautels Faculty of Management, McGill University*

July 31, 2006

Abstract

Several papers have questioned the ability of multifactor affine models to extract interest rate volatility from the cross-section of bond prices. These studies find that the conditional volatility implied by these models is very poorly or even negatively correlated with model-free volatility. We provide an in-depth investigation of the conditional volatility of monthly Treasury yields implied by three-factor affine models. We investigate different specifications of the price of risk and different specifications of volatility. For long maturities, the correlation between model-implied and EGARCH volatility estimates is approximately 82% for yield differences and 92% for yield levels. For short-maturity yields, the correlation varies between 58% and 71% for yield differences and between 62% and 76% for yield levels. The differences at short maturities are largely accounted for by the number of factors affecting volatility. A model-free measure of the level factor is highly correlated with EGARCH volatility as well as model-implied volatilities, which explains most of our findings. We conclude that multifactor affine models are much better at extracting time-series volatility from the cross-section of yields than argued in the literature. However, existing models have difficulty capturing volatility dynamics at the short end of the maturity spectrum, perhaps indicating some form of segmentation between long-maturity and short-maturity bonds. These results are robust to the choice of sample period, interpolation method and estimation method.

JEL Classification: G12

Keywords: term structure model; affine; conditional volatility; segmentation hypothesis; time series; cross section; EGARCH.

*Preliminary, comments welcome. Jacobs is also affiliated with CIRANO and CIREQ and wants to thank FQRSC, IFM^2 and SSHRC for financial support. Karoui was supported by a grant from IFM^2. Any remaining inadequacies are ours alone. Correspondence to: Kris Jacobs, Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 1G5; Tel: (514) 398-5025; Fax: (514) 398-3876; E-mail: kris.jacobs@mcgill.ca.

1 Introduction

The class of multifactor affine term structure models (ATSMs) has emerged as the workhorse in the fixed income literature, and a consensus has emerged in the literature that three-factor ATSMs are needed to successfully capture certain stylized facts of the term structure of interest rates.[1] Empirical implementations of these models often find that the term structure can be characterized in terms of the interest rate level, the slope of the term structure and term structure curvature. However, recently a number of papers have questioned the ability of multifactor ATSMs, and of three-factor ATSMs in particular, to capture some important aspects of term structure dynamics. Part of this recent criticism has been directed at the ability of multifactor ATSMs to model volatility. While these models are able to generate the hump-shaped pattern for unconditional volatility in both swap and Treasury markets (see Dai and Singleton (2000, 2003)), several studies have questioned their ability to model conditional volatility. Using swap data, Collin-Dufresne, Goldstein and Jones (2004) find that a popular and well-documented three-factor affine model implies volatility paths that are negatively correlated with the GARCH volatility estimates of weekly changes in the six-month rate. Andersen and Benzoni (2005) use intra-day Treasury data to document that realized yield volatility is unrelated to principal components extracted from the cross-section, which proxy for model-implied volatility.[2]

These findings question some of the most important building blocks of ATSMs, and in fact more fundamentally question the validity of a large class of arbitrage based term-structure models. It is a key implication of these models that the yields' conditional volatility is a linear combination of the state variables. For example, in the model used by Collin-Dufresne et al. (2004), in which volatility is driven by a single state variable, the failure of the model can be explained by the fact that the volatility factor will reflect the level of the yield curve, which may not be highly correlated with the time series volatility of the six-month rate. This feature is often referred to as unspanned stochastic volatility (USV), and it reflects a tension between the time series and the cross-sectional properties of the model.

These empirical findings have far-reaching practical implications, because if the yield curve fails to span volatility, fixed income volatility risk cannot be hedged by positions in the bond market alone. Consequently, any term structure model that relates the conditional volatility to the cross-section of yields will fail to capture the time variability of the conditional volatility. To resolve these problems, Collin-Dufresne and Goldstein (2002) and Collin-Dufresne et al. (2004) have proposed a new family of affine models, labeled USV models, in which the volatility does not affect the

[1] See Chen and Scott (1993) and Balduzzi et al. (1996) for early multifactor models. See Duffie and Kan (1996) for a full characterization and Dai and Singleton (2000) for a useful classification of multifactor ATSMs. Litterman and Scheinkman (1991) established that three-factor models are able to explain a large part of the variation in bond yields. See Duffee (2002) and Duarte (2004) for recent implementations of three-factor ATSMs.

[2] Collin-Dufresne and Goldstein (2002) document that the conditional volatility implied by term structure models is unrelated to implied volatility from interest rate options.

cross section of bonds yields. However, there is considerable disagreement about the empirical performance of these models (see for example Bikbov and Chernov (2004), Collin-Dufresne et al. (2004) and Thompson (2004)).

This paper examines the ability of affine models to simultaneously match the cross-sectional and time series properties of the term structure of monthly Treasury yields. More specifically, we examine the ability of heteroskedastic three-factor ATSMs to capture the conditional volatility for a large cross-section of Treasury yields. Because reconciling the time series and cross-sectional properties of the model critically depends on the mapping between the physical and risk-neutral model dynamics, we pay particular attention to the market price of risk, and we investigate three classes of models: completely affine models, essentially affine models (Duffee (2002)), and extended essentially affine models (Cheridito, Filipović and Kimmel (2005)). We follow the popular classification by Dai and Singleton (2000), which is very appropriate for volatility modeling. Dai and Singleton (2000) characterize four different three-factor models, dependent on how many factors affect the conditional volatility. For each specification of the price of risk, we investigate $\mathbb{A}_1(3)$, $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models.[3] Rather than using the volatility factor as a proxy for the conditional volatility of the short rate, we regress the exact model-implied conditional volatility for yields of different maturities on EGARCH volatility estimates for different maturities.

We document a sizeable positive correlation between model-implied conditional volatility and EGARCH volatility estimates. For long-maturity yields, the correlation is very robust across models and is approximately 82% for yield differences and 90% for yield levels. The correlation for the shortest-maturity yields is much more dependent on the model and varies between 58% and 71% for yield differences and between 62% and 76% for yield levels. The correlations generally increase with maturity. For short maturities, models with a higher number of volatility factors generate higher correlations, but the specification of the price of risk seems inconsequential. We also investigate the out-of-sample modeling of conditional volatility, and we find that the models do a better job of forecasting long-maturity yields than short-maturity yields.

We provide more insight for these findings by reporting a large and statistically significant correlation between EGARCH volatility estimates and the level factor. Estimated correlations are rather robust across models, but they are again highest for models with a higher number of volatility factors. The correlation of conditional volatility with slope and curvature is smaller but nonetheless significantly estimated for many models and maturities.

These findings lead to two important conclusions. First, Collin-Dufresne et al. (2004) and Andersen and Benzoni (2005) conclude that affine models fail in spanning the volatility because the conditional variance of yields is restricted to be a linear combination of yields. They interpret their negative results as a failure of existing multifactor models and conclude that different models are needed. Based on our evidence, we believe that the verdict for multifactor ATSMs is much more positive. Our findings clearly suggest that yields are informative about conditional volatility.

[3]An $\mathbb{A}_m(N)$ model is an N-factor model with m volatility factors.

Second, whereas three-factor ATSMs do not fail as spectacularly as suggested in other studies, our findings suggest that bond markets are incomplete and that the form of the incompleteness is a segmentation of the market. Yields with a maturity under one year have a very different volatility structure from long maturity yields, and the conditional volatility of short-maturity yields seems to exhibit an idiosyncratic variation that affine models are not able to match when calibrated on the full cross-section, regardless of the specification of the market price of risk. The reported tension between the cross-sectional and time series properties of the yield curve may therefore reflect some type of segmentation between the time series properties of the short and long end of the term structure.[4]

To provide some additional motivation for this interpretation, we separately estimate the short and long end of the term structure with three-factor models. In this case, the correlation from the regression of the model-implied volatility on the model-free volatility is roughly the same across maturities, which supports our interpretation. We document a number of other stylized facts suggesting that three-factor ATSMs have difficulty characterizing the conditional volatility at the short end of the yield curve. While the number of factors affecting volatility and the price of risk somewhat affect the ability of the model to match the correlation between yield volatilities and between conditional volatility and term structure factors at different maturities, it is a robust conclusion that the models experience their most serious problems at the short end of the maturity spectrum.

We investigate the robustness of our findings with respect to the choice of sample period, estimation method, and interpolation technique for the term structure data, and the results change very little. We also investigate the robustness with respect to the use of model-free benchmark by using different GARCH models, and by using instantaneous conditional volatility as an alternative to GARCH models, and again the results are robust. When we use realized volatility instead of EGARCH volatility as a measure of model-free volatility, the results change somewhat, but the correlation between model-free and model-implied volatility is still around 60%.

Even though Collin-Dufresne et al. (2004) and Andersen and Benzoni (2005) reach diametrically opposite conclusions, we note that some of our empirical results are confirmed by other studies. Dai and Singleton (2003) report that the $\mathbb{A}_1(3)$ model performs quite satisfactory when it comes to match the time variation in the conditional volatility of the 5-year yield in both swap and Treasury markets. Thompson (2004) documents a correlation of 56% between a short-maturity volatility LIBOR forecast and model-free GARCH volatility, which is similar to the correlation we find at short maturities. Almeida, Graveline and Joslin (2006) study swap rates and find correlations similar to ours for longer maturities, but lower correlations for shorter maturities. Bikbov and Chernov (2004) conduct a more ambitious comparison of Gaussian term structure models, stochastic

[4]The literature on segmentation between short-maturity and long-maturity fixed income securities goes back a long way (see for example Modigliani and Sutch (1967)). See Duffee (1996) for more recent evidence on idiosyncratic variation in short-maturity yields and segmentation in Treasury markets.

volatility models and USV models, using data on Eurodollar futures and options. While their focus is on the economic and statistical comparison of these models based on estimation using different data, their investigation confirms that affine term structure models are able to match time variation in the state variables, which is consistent with our findings.

The paper proceeds as follows. Section 2 discusses the modeling of conditional volatility in affine term structure models and the restrictions on volatility implied by the theory. Section 3 introduces the data and discusses the estimation technique. Section 4 presents the empirical results. Section 5 reports on an extensive robustness exercise and Section 6 concludes.

2 Conditional Volatility in Affine Term Structure Models

2.1 Affine Term Structure Models

We study affine models where the short rate is given by $r_t = \delta_0 + \delta_1 X_t$. The state vector X_t follows an affine diffusion under the risk-neutral measure $\mathcal{Q}$

$$dX_t = \widetilde{\kappa}\left(\widetilde{\theta} - X_t\right)dt + \Sigma\sqrt{S_t}d\widetilde{W}_t, \tag{1}$$

where $\widetilde{W}_t$ is a $N-$dimensional vector of independent standard $\mathcal{Q}-$Brownian motions, $\widetilde{\kappa}$ and Σ are $N \times N$ matrices and S_t is a diagonal matrix with a ith diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i' X_t.$$

From Duffie and Kan (1996), we know that

$$P(t,\tau) = \exp(A(\tau) - B(\tau)' X_t),$$

where $A(\tau)$ and $B(\tau)$ satisfy the following ODEs

$$\frac{dA(\tau)}{d\tau} = -\widetilde{\theta}'\widetilde{\kappa}B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma B(\tau)\right]_i^2 \alpha_i - \delta_0 \tag{2}$$

and

$$\frac{dB(\tau)}{d\tau} = -\widetilde{\kappa}B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma B(\tau)\right]_i^2 \beta_i + \delta_1 \tag{3}$$

Equations (2) and (3) can be solved numerically with the initial conditions $A(0) = 0$ and $B(0) = 0_{\mathbb{R}^N}$.[5] Throughout, we will use the classification scheme proposed by Dai and Singleton (2000)).

The model is completely specified through the dynamics of state prices. The pricing kernel π_t is given by

$$\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t' dW_t, \tag{4}$$

[5] We use the Rosenbrock method rather than the Runge-Kutta method because it is more convenient for stiff differential equations, which is sometimes the case for the $\mathbb{A}_1(3)$, $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models.

where W_t is a $N-$dimensional vector of independent standard $\mathcal{P}-$Brownian motions and Λ_t denotes the market price of risk. The dynamics of the state vector under the actual measure $\mathcal{P}$ can be obtained by subtracting $\Sigma\sqrt{S_t}\Lambda_t$ from the drift of equation (1).

2.2 The Price of Risk in Affine Term Structure Models

Several specifications for the market price of risk Λ_t are available in the literature. Traditionally the market price of risk was specified as

$$\Lambda_t = \sqrt{S_t}\lambda_0, \tag{5}$$

Models that adopt this specification are referred to as completely affine models. Duffee (2002) proposes a more general specification that leads to the essentially affine class of models (see also Dai and Singleton (2000,2002) and Duarte (2003))

$$\Lambda_t = \sqrt{S_t}\lambda_0 + \sqrt{S_t^-}\lambda_1 X_t, \tag{6}$$

where λ_0 is an $N \times 1$ vector of constants, λ_1 is a $N \times N$ matrix of constants and the diagonal matrix S_t^- has zeros in its first m entries and $\left(\alpha_i + \beta_i' X_t\right)^{-1}$ for $i = m+1, ..., N$. Note that the variance of the pricing kernel is not an affine function of the state vector, but since the latter does not affect bond prices, the affine property is maintained.

From (6), it can be seen that essentially affine models offer a number of advantages. They allow Λ_t (and the term premia) to vary independently of the level of the volatility and remove the restriction on its sign. Duffee (2002) shows that this improves the model's ability to match the time variability of excess returns.

Recently, Cheridito, Filipović and Kimmel (2005) extend the essentially affine class of models as follows

$$\Lambda_t = \sqrt{S_t}^{-1}\lambda_0 + \sqrt{S_t}^{-1}\lambda_1 X_t, \tag{7}$$

where λ_0 is an $N \times 1$ vector of constants, λ_1 is a $N \times N$ matrix of constants such that $\lambda_{1(ij)} = 0, \forall i \leq m$ and $j > m$. The resulting models are referred to as extended essentially affine models.

The literature has hitherto focused on the implications of the price of risk specification for modeling excess returns. When the market is complete, i.e. the dimension of the vector of Brownian motions is smaller than the number of traded bonds, the no-arbitrage condition combined with Itô's lemma implies that the bond-price dynamics can be written as

$$\frac{dP(t,\tau)}{P(t,\tau)} = \left(r_t + e_{\tau,t}\right) dt + V_{\tau,t} dW_t, \tag{8}$$

where the instantaneous expected excess return and its volatility term are restricted as follows

$$e_{\tau,t} = -B(\tau)'\Sigma\sqrt{S_t}\Lambda_t \text{ and } V_{\tau,t} = -B(\tau)'\Sigma\sqrt{S_t}. \tag{9}$$

The specification of the term premia or the expected excess return on a given maturity bond differs significantly between essentially affine and extended essentially affine specifications. From equation (9) , for the essentially affine model we have

$$e_{\tau,t} = -B(\tau)'\Sigma \left(S_t \lambda_0 + I^- \lambda_1 X_t \right),$$

where I^- is a diagonal matrix with elements equal to one except the first entry that is equal to zero. Although excess returns implied by this model can vary independently from the level of the conditional variance matrix S_t, the link between the conditional variance and the expected excess return is not totally broken. In contrast, the extended essentially affine model allows the term premia to be completely independent of the level of the conditional volatility. The form of the term premia in this case is

$$e_{\tau,t} = -B(\tau)'\Sigma \left(\lambda_0 + \lambda_1 X_t \right).$$

Note that this form is similar to the one implied by essentially affine Gaussian models. A key ingredient of the success of the $\mathbb{A}_0(3)$ (see Duffee (2002) and Dai and Singleton (2002)) may be the simple stylized fact that the term premia exclusively depend on the state vector. The extended essentially affine model specifies the market price of risk in a way that preserves the benefits of the Gaussian model, i.e. an affine term premium, while taking into account time varying conditional volatility.

It is important to consider alternative specifications for the price of risk for this study, because flexibility in modeling excess returns may extend to modeling conditional volatility. However, the additional flexibility of the extended essentially affine specification comes at a cost: the Feller condition has to be satisfied to rule out arbitrage opportunities, which makes extended essentially affine models more constrained than their essentially affine counterparts.[6] Whether the model's additional flexibility helps to match second moments of yields depends on the impact of the Feller condition on the parameter estimates.

While the time series performance of ATSMs depends on the specification of the market price of risk, their cross-sectional performance depends to a large extent on the number of factors. We limit our empirical investigation to three-factor models, because there is substantial evidence that (at least) three factors are needed to explain the variation in yield co-movements (see for example Litterman and Scheinkman (1991) and Dai and Singleton (2000)). Furthermore, for the purpose of modeling conditional volatility, it seems natural to use the classification of Dai and Singleton (2000), who organize three-factor models based on the number of factors impacting on the conditional volatility.[7] Since our goal is to assess the behavior of the conditional volatility in affine models, it seems natural to consider a heteroskedastic model, and therefore we only consider $\mathbb{A}_1(3)$, $\mathbb{A}_2(3)$ and

[6] The Feller condition implies that the volatility factors cannot attain the zero boundary. Since the market price of risk is proportional to the volatility factors, it will always be finite if the Feller condition is satisfied.

[7] Collin-Dufresne et al. (2004) use a somewhat different canonical representation of these models based on this classification in their empirical work.

$\mathbb{A}_3(3)$ models. While some other papers document the ability of the homoskedastic $\mathbb{A}_0(3)$ model to capture conditional volatility, we eliminate this rather contradictory model feature by using analytical conditional moments, rather than the reprojection technique of Gallant and Tauchen (1998). A formal description of the $\mathbb{A}_1(3)$, $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models under different specifications of the price of risk is provided in Appendix A.

2.3 Restrictions on Conditional Yield Volatility in Affine Models

In affine models, the conditional variance of a given maturity yield is an affine function of the state variable or equivalently a linear combination of yields. The conditional variance of a yield with maturity τ is

$$var_t\left(y_{t+h}\right) = \overline{B}'(\tau) var_t(X_{t+h}) \overline{B}(\tau), \tag{10}$$

where

$$\overline{B}(\tau) = \frac{B(\tau)}{\tau}. \tag{11}$$

Using the Kronecker product operator $\otimes$ and the fact that

$$vec\left(ABC\right) = \left(C' \otimes B\right) vec\left(A\right),$$

where vec denotes the vectorized representation of a matrix, we get[8]

$$\begin{aligned} var_t\left(y_{t+h}\right) &= \left(\overline{B}'(\tau) \otimes \overline{B}(\tau)\right) \times vec\left(var_t(X_{t+h})\right) \\ &= b_0 + b_1 X_t. \end{aligned} \tag{12}$$

The relationship in (12) is quite restrictive, because it indicates that in affine models the conditional yield variance is an affine function of the state vector. As such, it seems that the conditional variance is severely constrained in these models. In fact, (12) suggests that the conditional variance implied by affine models estimated using cross-sections of yields may be more intimately related to the properties of the level of yields than to the true volatility of these yields. Indeed, Collin-Dufresne et al. (2004) and Andersen and Benzoni (2005) have found that volatilities extracted from cross-sections of swap and Treasury yields are essentially unrelated to conventional volatility measures. In this sense the restriction in (12) has been interpreted as an indication of the failure of affine models in these markets.

It must be noted at this point that the ability of affine models to extract volatility from the cross-section can be evaluated in different ways. The empirical framework may differ dependent on whether one is exclusively interested in volatility or also in the modeling of the conditional mean. The approach used by Collin-Dufresne, Goldstein and Jones (2004) and Thompson (2004) is rather similar to ours, because they focus on volatility and evaluate the volatility measure of interest using an analytical formula, as we do with (12). Dai and Singleton (2003) and Bikbov and

[8] Note that while the restriction in (12) is on conditional variances, we empirically investigate correlations between model-free and model-implied volatility. Regressions using variances yield fairly similar correlations.

Chernov (2004) are interested in modeling the first two conditional moments, and therefore use a different approach based on the reprojection method (Gallant and Tauchen (1998)). This involves estimating a GARCH model on either the model-implied time series of yields, or on a simulated time series. The resulting estimate of volatility may be dependent on the specification of the model for the conditional mean.

A related issue is that some studies focus on the conditional volatility of the yields, while others investigate the conditional volatility of the instantaneous short rate. Our focus on (12) is practically motivated: the conditional yield volatility is available analytically for any maturity. Consequently a discretization of the short rate is not necessary to infer the level of conditional volatility.

3 Data and Estimation Technique

3.1 Data

For our main results, we use zero-coupon Treasury bond yields with maturities of 3 months, 6 months, 1 year, 2 years, 5 years and 10 years that are extracted using the unsmoothed Fama and Bliss (1987) method. Several other studies have used these data, see for example Ang and Piazzesi (2003), Cochrane and Piazzesi (2005) and Duffee (2005). Monthly observations for these data are available from January 1970 to December 2003.[9] We use data from 1970 to 1999 for in-sample estimation, and the period 2000-2003 is used for an out-of-sample exercise.

We conduct a robustness check to verify if the results depend on the sample size and the interpolation scheme. We use zero-coupon yields for maturities of 3 months, 6 months, 1 year, 2 years, 5 years and 10 years constructed using the McCulloch (1975) cubic spline interpolation method.[10] Monthly observations for these data are available from January 1952 to December 2003. Originally, this technique was used by McCulloch and Kwon (1991) to construct a dataset that ends in February 1991. Robert Bliss periodically updates this data set using a slightly different method (see Bliss (1997)). Although the yields constructed by McCulloch and Kwon (1991) and Bliss (1997) are somewhat different in overlapping periods, we use the McCulloch and Kwon (1991) data until February 1991 and the Bliss data thereafter. This dataset allows us to conduct two robustness exercises. We first compare the 1970-1999 results obtained using the cubic spline interpolation method with those obtained using the unsmoothed Fama and Bliss method. Subsequently we compare the results for the 1952-1999 sample with those for the 1970-1999 sample.

3.2 Estimation Technique

To estimate the models, we use the quasi-maximum likelihood (QML) method as implemented by Chen and Scott (1993) and Fisher and Gilles (1996). QML relies on the following state-space

[9] We thank Robert Bliss for graciously providing us with the data.

[10] This method is used among others by Duffee (2002), Duarte (2004) and Cheridito et al. (2006).

representation

$$y_t(\tau) = \overline{A}(\tau) + \overline{B}'(\tau)X_t, \tag{13}$$

and

$$dX_t = \kappa(\theta - X_t)\,dt + \Sigma\sqrt{S_t}dW_t, \tag{14}$$

where

$$\overline{A}(\tau) = -\frac{A(\tau)}{\tau} \text{ and } \overline{B}(\tau) = \frac{B(\tau)}{\tau}.$$

Although the state vector is not observed, it can be inverted from observed yields using equation (13) . However, because there are more observed yields than unobserved state variables, the market completeness hypothesis implies that yields other than those used to invert the state vector are perfectly predicted. In order to circumvent this problem, we use an approach that has become standard in the literature: we assume that we exactly observe a number of yields equal to the number of state variables and assume that the remaining yields are measured with error. To set the notation, assume that there are M observable yields, among which N are observed exactly and are denoted by the N dimensional vector $y_t(\tau^\star)$. The remaining $(M - N)$ are assumed to be measured with error and are denoted by the $(M - N)$ dimensional vector $y_t(\tau^\cdot)$. Equation (13) can then be written as

$$y_t(\tau^\star) = \overline{A}(\tau^\star) + \overline{B}'(\tau^\star)X_t \tag{15}$$

and

$$y_t(\tau^\cdot) = \overline{A}(\tau^\cdot) + \overline{B}'(\tau^\cdot)X_t + u_t, \tag{16}$$

where

$$u_t \sim \mathcal{N}(0, \Sigma) \text{ and } \Sigma = L'L. \tag{17}$$

The conditional probability distribution of yields is obtained from the conditional distribution of the state vector via the Jacobian of the affine transformation (15) that relates yields to the state vector and the distribution of the measurement errors. For a given date t, the conditional probability density function can be written as

$$f(y_{t+1}(\tau) \mid y_t(\tau)) = \frac{1}{|\det(\mathbb{B})|} f\left(X_{t+1} \mid \widehat{X}_t\right) + g(u_{t+h}), \tag{18}$$

where $\widehat{X}_t$ is the implied value of the state vector computed as

$$\widehat{X}_t = \mathbb{B}^{-1}\left(y_t(\tau^\star) - \overline{A}(\tau^\star)\right), \tag{19}$$

and

$$\mathbb{B} = \begin{bmatrix} \overline{B}'(\tau_1^\star) \\ \cdot \\ \cdot \\ \overline{B}'(\tau_N^\star) \end{bmatrix}. \tag{20}$$

QML is then relatively easy to implement since it only relies on the two first conditional moments of the state vector and does not impose any restriction on the parameters of the model. The log likelihood of the t^{th} observation is

$$\begin{aligned}\mathcal{L}_t(\Theta) &= -\frac{M}{2}\log(2\pi) - \log(|\det(\mathbb{B})|) - \frac{1}{2}\log\left(\det\left(L'L\right)\right) - \frac{1}{2}\log\left(\det\left(var_{t-1}(X_t)\right)\right) \quad (21)\\ &\quad -\frac{1}{2}(\widehat{X}_t - E_{t-1}(X_t))' var_{t-1}(X_t)(\widehat{X}_t - E_{t-1}(X_t)) - \frac{1}{2}u_t^{\cdot\prime}\left(L'L\right)^{-1}u_t^{\cdot},\end{aligned}$$

where the measurement error vector $u_t^{\cdot}$ is obtained from (16).

In optimization, the initial log likelihood is computed using the unconditional distribution of the state vector. We assume that the total number of observed yields is equal to six. We assume that the 6-month, 2-year and 10-year yields are observed exactly, whereas the 3-month, 1-year and 5-year yields are assumed to be measured with errors. The expressions for the two first conditional moments of the state vector are provided in Appendix B.

4 Empirical Results

4.1 Parameter Estimates

We estimate term structure models for the in-sample period January 1970-December 1999. Maximizing the log likelihood function for models of this kind is not straightforward and it is critical to find good starting values for the parameters to ensure convergence to a global optimum and avoid non-admissible dynamics (see Duffee (2002) for more on convergence problems in affine term structure models). We use the following heuristics to improve the optimization procedure. We first simulate 10000 admissible starting values using the multivariate normal distribution based on plausible means and variances. By admissible, we mean starting values that ensure the positivity of the state variables and satisfy the admissibility and stationarity conditions described in the appendix. For each of these 10000 starting values, we compute the log likelihood function and choose the 50 best starting values. Optimization is then performed 50 times using these starting values and subsequently we pick the parameter vector that maximizes the likelihood function among these local optima. It is reassuring that this optimal parameter vector is obtained repeatedly with different starting values.

Tables 1-3 report the parameter estimates, which are roughly consistent with the available literature. Note that Table 2 does not report estimation results for the essentially affine $\mathbb{A}_3(3)$ model, because the model is identical to the completely affine $\mathbb{A}_3(3)$ model. We also do not report estimation results for the extended essentially affine $\mathbb{A}_2(3)$ model in Table 3, because we were unable to obtain economically meaningful estimation results. Some of the parameter estimates for this model in Cheridito et al. (2005) are also hard to interpret. Because our investigation does not critically depend on this particular model, we decided it was safer to drop it.

The parameters satisfy admissibility and stationarity conditions under the actual measure $\mathcal{P}$

for all models.[11] As is now well established in the literature, the state variables can be thought as proxies for the level, slope and curvature factors identified by Litterman and Scheinkman (1991). We verified this by examining the response to shocks to each of the state variables. The time series properties of the state vector critically depend on the matrix with the speed of mean reversion κ. For the $\mathbb{A}_1(3)$ model, the first state variable X_{1t} that drives the conditional volatility is very persistent, as demonstrated by the first column of the matrix κ. The second state variable is as persistent as the volatility factor and the third state variable reverts quickly to its mean. Under the essentially and extended essentially affine specifications, the estimates of the vector λ_0 and the matrix λ_1 show that the market price of risk and the instantaneous excess return on bonds are mainly driven by the second and the third state variable. This suggests that the volatility factor X_{1t} plays a minor role in capturing the time variability of the instantaneous excess returns, and confirms that both essentially and extended essentially affine models free up the term premia from the level of the conditional volatility. For the $\mathbb{A}_2(3)$ model, the first state variable X_{1t} is the most persistent and corresponds therefore to the level factor, whereas the second state variable X_{2t} exhibits slightly less persistence and the third state variable is the least persistent. Interestingly, the completely and essentially affine $\mathbb{A}_2(3)$ models are quite comparable as the estimates of the matrix λ_1 are close to zero for the essentially affine specification. This result is consistent with the findings in Duffee (2002), who uses a different data set. For both completely and extended essentially affine $\mathbb{A}_3(3)$ models, the first state variable becomes a curvature factor, the second state variable can be viewed as a level factor, whereas the third variable represents the slope factor. Once again, this can be seen by examining the response to shocks to each of these state variables.

4.2 Variance Forecasts

We evaluate the quality of a model's variance forecast by comparing the model-implied conditional variance of yields with the "true" conditional variance implied by an EGARCH(1,1) model.[12] The model-implied conditional variance of yields is given by

$$var_t(y_{t+1}(\tau)) = \frac{B(\tau)'}{\tau} var(X_{t+1} \mid \widehat{X}_t) \frac{B(\tau)}{\tau} + var(u_{t+1}(\tau)), \tag{22}$$

where $\widehat{X}_t$ is the inverted value of the state vector, and $u_{t+1}(\tau)$ is the measurement error (if the yield is not measured exactly) that corresponds to the yield with maturity τ. Much of our empirical investigation uses the conditional variance of yield differences, which is also equal to (22), but is compared with the EGARCH volatility estimates for yield differences.

To assess the quality of the forecast, we compute the unconditional correlation between the model-implied volatility forecast and the EGARCH(1,1) conditional volatility. Several studies have computed the model-implied conditional variance of yields using the reprojection technique of

[11] Note that we do not impose stationarity under the risk-neutral measure $\mathcal{Q}$ since this condition is not critical for estimation.

[12] Results were very similar when we used a GARCH(1,1) model instead of an EGARCH(1,1) model.

Gallant and Tauchen (1998) rather than by using (22). This technique is typically used when evaluating the first two conditional moments. The basic idea behind this approach is to assess the performance of a model by comparing the conditional probability distribution implied by the data to that implied by the model. This is implemented by extracting the model implied time series of the unobserved factors, computing the in-sample yield forecasts and subsequently estimating a GARCH model on these forecasts. The estimates of the GARCH model are then compared to the GARCH model estimated on the actual data. For studying conditional volatility, we prefer using (22) because it is simpler and because other authors have noted that in some cases the results of the reprojection technique may be hard to interpret. For example, Bikbov and Chernov (2004) apply this technique using Eurodollars futures and options data and find that the homoskedastic $\mathbb{A}_0(3)$ model does a good job in fitting the conditional volatility. As noted by Bikbov and Chernov (2004), this is a surprising result. A potential explanation is that the state vector and the errors are either filtered or inverted from observed yields. For instance, if QML is used, the state vector is usually inverted from yields that are observed without errors and are therefore a linear combination of yields with various maturities. As a result, the heteroskedasticity of the yields will show up in the time series of the implied state vector even though the original dynamics of the state vector are homoskedastic. In our approach using (22), the state vector is inverted from the observed yields, but the forecast of the variance is constructed using the statistical properties of the model, which eliminates the paradoxical possibility that a homoskedastic model is considered as a good model for fitting conditional volatility.

Tables 4 and 5 report on regressions of model-implied volatility on EGARCH estimates of the first difference and the level of yields with an AR(1) specification for the conditional mean.[13] Following the literature, we report correlation coefficients rather than R-squares. Panel A reports these correlation coefficients for the different term structure models. The R-square can be obtained as the square of the correlation coefficient. For future reference, Panels B and C report more detailed results for two models, the essentially affine $\mathbb{A}_1(3)$ model and the extended essentially affine $\mathbb{A}_3(3)$ model, which were chosen because their empirical performance is somewhat different. Standard errors are computed using the Newey and West (1987) estimator.

Table 4 documents a robust and significant positive relationship between model-implied and EGARCH volatility estimates of changes in yields at all maturities. We document a sizeable positive correlation between model-implied conditional volatility and EGARCH volatility. The correlation for long-maturity yields is very robust across models and is approximately 82%. The correlation for the shortest-maturity yields is much lower, much more dependent on the model and varies between 58% and 71%. Table 5 indicates that the correlation is even higher when EGARCH volatility estimates on yields are used as a model-free benchmark. The correlation is approximately

[13] We investigated the robustness of the results with respect to alternative specifications of the conditional mean. An AR(2) and an ARMA(1,1) specification yield virtually the same results. A constant conditional mean implies an even higher correlation for the first difference in yields and a lower correlation for the level of yields, but we found substantial evidence that this model is misspecified.

92% at the long end of the yield curve and varies between 62% and 76% at the short end. These results are clearly in contrast with the small positive or even negative correlation reported elsewhere in the literature (see Collin-Dufresne et al. (2004) and Andersen and Benzoni (2005)). Figures 1 and 2 further emphasize these findings by plotting the conditional volatility implied by the model and the EGARCH estimates of changes in yields over time for two cases, the essentially affine $\mathbb{A}_1(3)$ model and the extended essentially affine $\mathbb{A}_3(3)$ model. While the model-implied volatility is too smooth compared to the EGARCH estimates for most maturities, the problem is exacerbated at short maturities. Figure 3 provides additional intuition on the differences between the short and the long end of curve. This figure depicts the autocorrelation path of the standardized squared changes in yields implied by the essentially affine $\mathbb{A}_1(3)$ model, and the Bartlett standard errors bands.[14] The standardized squared change in yields is computed as

$$\frac{\Delta y_t^2\left(\tau\right)}{\sqrt{var_t(y_{t+1}\left(\tau\right))}}, \tag{23}$$

where $var_t(y_{t+1}\left(\tau\right))$ is computed as (22). A good variance model should imply an autocorrelation path that has no systematic patterns. Clearly, one can see that this goal is only achieved for long maturity yields (5 and 10-year); the shorter the maturity, the poorer is the performance of the model.

Tables 4 and 5 suggest that while affine term structure models do a much better job of extracting volatility from the cross-section of yields than suggested in some of the available literature, even the most complex models we consider have substantial shortcomings. We now provide some additional evidence on these shortcomings, with an emphasis on the differences between the modeling of the volatility of short maturity and long maturity yields. To save space, we only report on results for yield differences. Results for the levels of yields are very similar.

We first focus on the correlation between volatilities for yield differences of different maturities. Panel A in Table 6 presents the correlation matrix for EGARCH volatilities at different maturities. The correlation is close to one for yields that are nearer in maturity, but much lower between short-maturity and long-maturity yields. Panels B-E present correlation matrices for model-implied volatilities. Note that by definition, these correlations are equal to one for $\mathbb{A}_1(3)$ models. Panels B and C in Table 7 indicate that $\mathbb{A}_2(3)$ models cannot capture the correlation between volatilities for short-maturity and long-maturity bonds. Panels D and E indicate that $\mathbb{A}_3(3)$ models perform much better in this dimension but still fall short of the EGARCH-based correlations in Panel A. To demonstrate how poor the performance of most models is in this dimension, consider Table 7 which reports similar correlation matrices for yield levels for two very different models. For both models, model-implied correlations are close to the correlations obtained from the raw data.

[14] Results are very similar for the other models. The standard error bands correspond to plus/minus two standard errors arround zero.

4.3 Volatility and the Cross-Section of Yields

Our analysis demonstrates that a large proportion of the time variation in the conditional volatility can be inferred from the cross-section of yields, using the model-implied variance of the state variables in (22). Note that because the model-implied variance is affine in the yields, an alternative approach would be to relate the time variation in conditional volatility more directly to the cross-section of yields, as in (12). Table 8 takes this perspective and reports correlations between model-free measures of the factors that capture variation in the yield curve, and either the EGARCH volatility or the model-implied volatility. Based on the results of Litterman and Scheinkman (1991), we investigate level, slope and curvature factors. The level factor is proxied by the 3-month yield, the slope factor is measured by the difference between the 10-year and the 3-month yields whereas the curvature is calculated as the yield on a butterfly that is long on the 10-year and 3-month bonds and short on the 1-year bond. This approach is similar to Collin-Dufresne et al. (2004). Andersen and Benzoni (2005) use a related approach, in which they use principal component analysis to reduce the dimensionality of the cross-section of yields.

Table 8 indicates a high positive correlation between EGARCH volatility and the level of the yield curve. The correlations range between 66% and 77%, and are not affected by the maturity of the bond. Correlations between EGARCH volatility and slope and curvature are usually negative and, in contrast with the level, are model- and maturity-dependent. Whereas the $\mathbb{A}_1(3)$ model is able to capture the co-movements between volatility and the level of the yields, it does not capture the co-movements between volatility and the slope of the yield curve nor the co-movements between volatility and the curvature of the yield curve very well. The reason is that the volatility in this one-factor volatility model is simply a proxy for the level factor. The $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models capture the correlation between the volatility and all the factors affecting the yield co-movements much better. Presumably this is due to the fact that more factors drive the volatility. Interestingly, differences between EGARCH-based correlations and correlations implied by the $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models are less prominent at the long end of the curve.

These results are entirely consistent with the regression results in Tables 4 and 5. The correlations between model-free (EGARCH) and model-implied volatility are higher for the $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models, because these models are able to incorporate the component of volatility that is accounted for by the slope and curvature factors. The $\mathbb{A}_1(3)$ model on the other hand only picks up the volatility correlated with the level factor, but because this factor accounts for a very large percentage of the cross-section, the differences with the $\mathbb{A}_2(3)$ and $\mathbb{A}_3(3)$ models in Tables 4 and 5 are not very large. Our findings contrast with those of Collin-Dufresne et al. (2004), who document a small unconditional correlation between the EGARCH volatility estimates and level, slope, and curvature using the 6-month LIBOR rate. This finding may indicate a fundamental difference between swap and Treasury markets.

4.4 Volatility and the Market Price of Risk

Our findings show that the specification of the price of risk has a very minimal impact on the results. Dai and Singleton (2003) report that an essentially affine $\mathbb{A}_1(3)$ model substantially outperforms its completely affine counterpart for the purpose of volatility modeling, which contrasts with our findings. We replicated our empirical exercise with the reprojection method used in Dai and Singleton (2003), and in this case our results are in line with theirs. We focus on the results in Tables 4 and 5 because we believe that the more direct approach in (12) is preferable when focusing on conditional volatility. The reprojection technique is more appropriate when assessing the ability of the model to jointly capture the time varying nature of excess returns and conditional volatility.

There is no a priori reason that renders the market price of risk irrelevant for the extraction of conditional volatility. A closer examination of the conditional volatility as implied by equation (22) provides the intuition behind this surprising empirical result. The conditional volatility implied by affine models depends on two components: the factor loadings $B(\tau)$ and the conditional variance of the state vector. It turns out that the factor loadings, which determine the cross-sectional performance of the model, are almost identical across models. Moreover, the conditional variance of the state vector is determined by the components of κ and β that drive the conditional volatility, and those estimates are roughly the same across specifications of the price of risk.

4.5 The Segmentation Hypothesis

The verdict on the ability of affine models to extract conditional volatility from the cross-section of yields is decidedly mixed. Clearly, for our data and using the EGARCH measure for true volatility, the performance of affine models is not nearly as bad as suggested in several places in the literature. Nevertheless, Figures 1-2 clearly indicate that model-implied volatility forecasts are too smooth over time, and Table 8 indicates that the model-implied correlation with the state variables is very different from the correlation with the EGARCH measure. In this Section we further investigate the finding that the model seems to match the data much better for longer maturities. To save space, we report results on the essentially affine $\mathbb{A}_1(3)$ model. The results for the other three-factor models are similar.

The interest rate and term structure literature has repeatedly observed that the dynamics of short maturity yields are different from those of long maturity yields (see for example Duffee (1996) and Modigliani and Sutch (1976)). In its most extreme form, this hypothesis is known as the segmentation hypothesis, indicating that the short and long end of the term structure are segmented. We investigate the relevance of this hypothesis for estimating conditional volatility by repeating the analysis using only yields with maturities less than one year in estimation. More precisely, we use zero-coupon bond yields with maturities of 1 month, 3 months, 5 months, 7 months, 9 months and 12 months that are extracted using the unsmoothed Fama and Bliss method. We also estimate the model using only yields with maturities of more than one year in estimation, namely zero-coupon bond yields with maturities of 1 year, 2 years, 3 years, 5 years, 7 years and 10 years.

Table 9 presents the parameter estimates for the essentially affine $\mathbb{A}_1(3)$ model. Clearly there are some important differences between the parameter estimates for short and long maturities, as well as with the estimates in Table 2. Comparing the estimates of the matrix λ_1 with the estimates in Table 2, one can see that the curvature factor impacts more on the term premia. This is not really surprising, as such a shock can indeed be viewed as a "flight to quality" shock that primarily affects the willingness of investors to hold short-maturity bonds.[15] Perhaps a more important difference between the estimates of Tables 9 and 2, at least for the purpose of volatility analysis, is the time series behavior of the state variables. The estimates of the matrix κ demonstrate that the state variables are overall more persistent than those implied by an estimation of the same model on the whole yield curve, which has implications for the volatility factor. A more persistent volatility factor could generate a more persistent conditional volatility and thus enhance the performance of the model at the short end of the curve. Table 10 presents the results from regressions of EGARCH volatility estimates of the first difference in yields on model-implied estimates. The correlations for the long-maturity yields in Panel B are lower than those in Table 2, but the ones for the short-maturity yields in Panel A are significantly higher.

In summary, there is evidence of segmentation between the markets for long-maturity and short-maturity bonds. The model performs much better in key dimensions when it is estimated separately for short-maturity and long-maturity bonds.

It is interesting to note that several studies document problems at the short end of the swap curve.[16] This is often attributed to the fact that swap data are not available for three and six month maturities and that LIBOR rates are used to complete the term structure data, which may lead to biases in estimation. It is an open question whether these results and results on Treasury markets such as ours are indicative of some more structural underlying segmentation between the short and long end of the fixed-income markets, or that the reason for the segmentation lies in institutional issues that are quite different in both markets.

5 Robustness Analysis

This Section reports a number of robustness exercises. To economize on space, we only report results for the essentially affine $\mathbb{A}_1(3)$ model, and for yield differences. Results for other cases are very similar.

5.1 Out-of-Sample Analysis

Table 11 verifies whether the in-sample differences in fit for different maturities also hold out-of-sample. To evaluate the model's performance, we compute the root mean squared error (RMSE),

[15] See Duffee (2002) for a more detailed discussion.

[16] See for example Liu, Longstaff and Mandell (2006) for swap rates and Piazzesi (2001) for Treasury yields.

which is based on the differences between model-implied conditional volatility and EGARCH volatility. Out-of sample EGARCH volatility estimates are obtained by extending the sample one month at a time and re-estimating an EGARCH(1,1) on the new extended sample. Out-of-sample conditional volatility implied by the essentially affine $\mathbb{A}_1(3)$ is computed using the parameters in Table 2. The results show that the in-sample difference between short and long-maturity yields also hold out-of-sample for both changes and levels of yields. This finding is very compelling evidence that the differences in fit between maturities are very robust, because the in-sample and out-of-sample data are very different. Most of the RMSE in the in-sample period 1970-1999 is driven by errors in 1973-1974 following the first oil crisis and the period following the monetary policy experiment in 1980-1982. In contrast, the out-of-sample period 2000-2003 is a very quiet period for Treasury markets.

5.2 An Analysis of Instantaneous Conditional Volatility

Instead of focusing on model-implied conditional yield volatility at different horizons, Collin-Dufresne et al. (2004) focus on the instantaneous volatility of the short rate implied by the model. Table 12 presents the results of their approach for our data. We regress the conditional volatility implied by EGARCH estimates of changes in yields based on 1-month and 3-month yields on the instantaneous volatility of the short rate implied by the essentially affine $\mathbb{A}_1(3)$ model. Panel A of Table 12 shows that the correlations are of the same order of magnitude as the short-maturity correlations in Table 2. Panel B shows that when re-estimating the model using short-maturity yields only, the correlations increase. These results are more directly comparable to those of Collin-Dufresne et al. (2004), who find a negative correlation using swap data.

5.3 An Analysis of Realized Volatility

While the use of a GARCH or EGARCH model to measure "true" conditional volatility over different horizons is a well-established technique in the literature on term-structure volatility,[17] it obviously has some drawbacks. The use of realized volatility is an interesting alternative.[18] Unfortunately, it is not possible to construct indicators for realized volatility using high-frequency data for our sample period. Therefore, we follow the technique pioneered by Schwert (1989) in the equity return literature and construct measures of monthly volatility using within-month squared changes in yields. Assuming that m observations are available within a month, an estimate of the monthly variance can be calculated as

$$\sigma^2_{m,t+1}(\tau) = \sum_{i=1}^{m} \Delta y^2_{t+i/m}(\tau). \tag{24}$$

[17] It is used among others by Bikbov and Chernov (2004), Collin-Dufresne et al (2004) and Dai and Singleton (2003).

[18] See Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001) for applications of realized volatility.

It has been shown that an ARMA(1,1) provides a good fit to the logarithm of the realized variance:

$$\log\left(\sigma_{m,t+1}^{2}\left(\tau\right)\right)=\alpha\log\left(\sigma_{m,t}^{2}\left(\tau\right)\right)+\delta\varepsilon_{t}+\varepsilon_{t+1}, \tag{25}$$

with $\varepsilon_{t+1}\sim\mathcal{N}\left(0,\sigma_{t,\varepsilon}^{2}\right)$.[19] The variance forecast implied by this model is

$$\sigma_{m,t+1|t}^{2}\left(\tau\right)=\left(\sigma_{m,t}^{2}\left(\tau\right)\right)^{\alpha}\exp\left(\delta\varepsilon_{t}+\frac{\sigma_{t,\varepsilon}^{2}}{2}\right) \tag{26}$$

Panel A of Table 13 reports the correlation coefficients between the volatility forecast (26) and the $\mathbb{A}_1(3)$ model-implied volatility. The correlation between the essentially affine $\mathbb{A}_1(3)$ model-implied volatility and the realized volatility forecast is positive and significant for all maturities. However, at long horizons the correlations are lower than the corresponding ones in Table 4. Interestingly, Panel A of Table 13 indicates that at the long end of the curve the EGARCH estimates are also not very highly correlated with the realized volatility estimates. Figure 4 depicts the differences between the conditional volatility implied by the realized variance model and that implied by the EGARCH model, and it can be seen that the differences mostly occur for longer maturities. A comparison of Figure 4 with the model-implied volatility in Figure 1 provides further intuition for these results.

What drives the differences between EGARCH conditional volatility and realized volatility evident from Figure 4? To provide more insight, Figure 5 presents the term structure of unconditional yield volatility implied by the monthly data, the essentially affine $\mathbb{A}_1(3)$ model, the EGARCH model and the realized variance model. The "monthly data" term structure of unconditional yield volatility is given by the standard deviation of the monthly changes in yields. The unconditional model volatilities are computed as the averages of the conditional volatility path generated by each model. Figure 5 shows that while there is a close correspondence between the "monthly data" term structure, the EGARCH term structure and the term structure implied by the unconditional essentially affine $\mathbb{A}_1(3)$ model, the term structure of unconditional realized volatility is very different.[20] It is tempting to interpret this as a failure of the realized volatility approach, but it must be kept in mind that the "monthly data" term structure is also just another measure of the true volatility. These results therefore merely reflect some interesting differences between volatility measures that deserve further study, but this is beyond the scope of this paper. Moreover, most recent studies have used high-frequency data to construct realized volatility, and the properties of this time series may differ significantly from the realized volatility data in Figure 5.

The realized volatility analysis in Panel A of Table 13 provides an analysis of ex-ante restrictions implied by the model, in line with the analysis in Table 4 and the rest of the paper. Andersen and Benzoni (2005) use realized volatility data and mostly focus on ex-post restrictions implied

[19] We also investigated an ARMA(1,1) with an EGARCH innovation, and this does not affect the results.

[20] We also explored the residuals of the ARMA(1,1) realized variance model. For long-maturity yields (5 and 10-year), the residuals contain more noise than short maturity yields.

by term structure models. A detailed comparison with their results is beyond the scope of this paper.[21] However, to provide a worst-case scenario for the correlation between model-implied and data volatility using our data, consider Panel B of Table 13. We compute the correlation between the model-implied variance and the (ex-post) realized variance, rather than the (ex-ante) conditional expectation of the realized variance as in Panel A. We do not regard this comparison as meaningful, because it involves an ex-ante and an ex-post measure, but the resulting correlations present a worst-case scenario precisely because we compare the volatile ex-post realized volatility measure with the smoother ex-ante model-implied measure. Panel B indicates that even in this case, the correlations are 50% on average.

We conclude that whereas there are some interesting differences between the results obtained using realized volatility and those obtained using EGARCH volatility, most notably at long maturities, the correlations between model-implied and "true" volatility are reliably positive, even if we use ex-post realized volatility as a measure of true volatility.

5.4 The Structure of the Measurement Errors

In this Section, we verify whether assumptions regarding measurement errors have a significant impact on our results. To this end, we first estimate the essentially affine $\mathbb{A}_1(3)$ model with QML under different assumptions on the measurement errors. We use zero-coupon bond yields with maturities of 6 months, 1 year, 2 years, 5 years, 7 and 10 years that are extracted using the unsmoothed Fama and Bliss method. We assume that the 6-month, 2-year and 7-year yields are observed without errors, whereas the 1-year, 5-year and 10-year yields are assumed to be measured with errors. As a second robustness check, we estimate the essentially affine $\mathbb{A}_1(3)$ model with the data set used for Table 2, but using the extended Kalman Filter combined with quasi-maximum likelihood, thereby allowing all yields to be measured with errors.[22] Panel A of Table 14 shows that even when different points on the yield curve are used to estimate the model, and different points on the yield curve are assumed to be observed exactly, the differences between the short and the long end of the yield curve prevail. Panel B of Table 14 presents results obtained using the extended Kalman filter. Compared to Table 4, the correlations increase slightly, but the differences in fit across maturities are very robust.

5.5 Sample Period and Interpolation Method

Comparisons between different models in the term structure literature are sometimes difficult because different techniques are used to interpolate zero-coupon bonds from raw yield data.[23] A large

[21] Andersen and Benzoni (2005) use Treasury data, but because they use measures of realized bond market volatility that are constructed using high frequency data, they use a much shorter sample for 1991-2001. The ex-ante test results they report are also very different from ours.

[22] See Hamilton (1994) for a general presentation and Duan and Simonato (1999) for the estimation of term structure models using the Kalman Filter.

[23] See Dai, Singleton and Yang (2003) for a discussion on the importance of the interpolation method.

number of papers use the unsmoothed Fama and Bliss method that we have used in our empirical work. We now investigate the robustness of our findings by re-estimating the model using zero-coupon yields interpolated using the equally popular McCulloch (1975) cubic spline approach. See Section 3.1 for details of our implementation of the procedure. Panel A of Table 15 presents the regression results using 1970-1999 data. Clearly the regression estimates and the correlations in Panel A are similar to those in Table 4. We therefore conclude that our results are robust to the choice of interpolation method.

As discussed in Section 3.1, the McCulloch-Kwon zero-coupon bond yields are available for the more extended time period 1952-1999. Panel B of Table 15 reports the results of the same regression using this extended sample. Results are similar, and we conclude that our results are robust with respect to the choice of sample period.

6 Concluding Comments

This paper conducts an investigation of the ability of three-factor affine term structure models to extract the time series of conditional yield volatility from the cross-section of Treasury yields. Several papers have documented that these models fail in this dimension, but we find evidence to the contrary. We find that for yield differences, the correlation between the conditional volatility of model-implied yields and model-free conditional volatility as measured by EGARCH volatility is approximately 82% at long maturities, and between 58% and 71% at short maturities. For yield levels, correlations are even higher. Models in which more factors impact on the conditional volatility perform better, but the specification of the price of risk impacts very little on results. We also provide evidence that the models perform worse for short maturities in several other dimensions, perhaps indicating some form of segmentation between the short and long end of the yield curve. To provide intuition for these findings, we document a high and robust correlation between EGARCH estimates and the level factor.

The empirical results are robust to a large number of variations in the empirical setup. When using realized volatility as a measure of volatility, results for long maturities are somewhat different, but correlations are still robustly positive. Our findings on the role of the price of risk specification are different from existing studies. The reason is that we use an analytical expression for model-implied volatility. Other studies do not exclusively focus on the second conditional moments of yields, and therefore use different techniques.

We do not provide a *solution* that enables three-factor ATSMs to capture the conditional volatility at the short end of the term structure, when the model is calibrated on the full cross-section. It is tempting to suggest that the models considered in this paper are not rich enough for the task at hand, and that a fourth factor may fix the problem. However, our results for three-factor models indicate that the four-factor models most likely to be successful will likely be in the $\mathbb{A}_3\,(4)$ or $\mathbb{A}_4\,(4)$ class, and these are very heavily parameterized models that are hard to estimate. Indeed, we speculate that the richness of the parametrization of such models may lead to a deterioration of the out-of-sample performance for many issues of interest, such as yield forecasting. We believe that the use of additional data such as option contracts to calibrate existing three-factor models may hold more promise. See among others Almeida, Graveline and Joslin (2006), Bikbov and Chernov (2004), Fan, Gupta and Ritchken (2003), Jagannathan, Kaplin and Sun (2003) and Li and Zhao (2005).

One caveat is that some related studies use different data, which may explain some of the differences in the results. Collin-Dufresne et al. (2004) use swap data. Almeida, Graveline and Joslin (2006) report results that are somewhat similar to ours for weekly swap rates. However, they find that correlations for short maturities are lower than ours. It may prove interesting to investigate if these differences are due to institutional differences between swap and Treasury markets, or perhaps to some other feature such as the frequency of the data. In general, an investigation of the robustness of our results with respect to data frequency may prove worthwhile.

Appendices

A Three-Factor ATSMs

The $\mathbb{A}_1(3)$ model has the following canonical representation under the physical measure $\mathcal{P}$

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{pmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{1+\beta_{21}X_{1t}} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_{1t}} \end{bmatrix} dW_t. \tag{27}$$

Admissibility conditions require that $\kappa_{11}\theta_1 \geq 0, \theta_1 \geq 0, \delta_{1(1)} \geq 0, \beta_{21} \geq 0$ and $\beta_{31} \geq 0$. The Feller condition requires $\kappa_{11}\theta_1 \geq \frac{1}{2}$. The form of the market price of risk used in the essentially affine $\mathbb{A}_1(3)$ model is

$$\Lambda_t = \begin{bmatrix} \lambda_{0(1)}\sqrt{X_{1t}} \\ \lambda_{0(2)}\sqrt{1+\beta_{21}X_{1t}} \\ \lambda_{0(3)}\sqrt{1+\beta_{31}X_{1t}} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\lambda_{1(21)}X_{1t}+\lambda_{1(22)}X_{2t}+\lambda_{1(23)}X_{3t}}{\sqrt{1+\beta_{21}X_{1t}}} \\ \frac{\lambda_{1(31)}X_{1t}+\lambda_{1(32)}X_{2t}+\lambda_{1(33)}X_{3t}}{\sqrt{1+\beta_{31}X_{1t}}} \end{bmatrix}, \tag{28}$$

or equivalently

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)}X_{1t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)}+\left(\lambda_{1(21)}+\lambda_{0(2)}\beta_{21}\right)X_{1t}+\lambda_{1(22)}X_{2t}+\lambda_{1(23)}X_{3t}}{\sqrt{1+\beta_{21}X_{1t}}} \\ \frac{\lambda_{0(3)}+\left(\lambda_{1(31)}+\lambda_{0(3)}\beta_{31}\right)X_{1t}+\lambda_{1(32)}X_{2t}+\lambda_{1(33)}X_{3t}}{\sqrt{1+\beta_{31}X_{1t}}} \end{bmatrix}. \tag{29}$$

For the $\mathbb{A}_1(3)$ model, the extended essentially affine model specifies Λ_t as follows

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)}+\lambda_{1(11)}X_{1t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)}+\lambda_{1(21)}X_{1t}+\lambda_{1(22)}X_{2t}+\lambda_{1(23)}X_{3t}}{\sqrt{1+\beta_{21}X_{1t}}} \\ \frac{\lambda_{0(3)}+\lambda_{1(31)}X_{1t}+\lambda_{1(32)}X_{2t}+\lambda_{1(33)}X_{3t}}{\sqrt{1+\beta_{31}X_{1t}}} \end{bmatrix}. \tag{30}$$

Clearly, the essentially affine model is nested by its extended counterpart only if the β coefficients are equal to zero. In this case, it imposes the restriction that $\lambda_{0(1)} = 0$. In other cases, the classical likelihood ratio test cannot be performed. The extended model also differentiates itself from the essentially affine model in that the Feller or the unattainability condition has to be satisfied to rule out arbitrage. For the purpose of estimation, we make the following change of variable to guarantee that the Feller condition is not violated

$$(\kappa\theta)_1 = \left(\kappa\theta_1^{*2} + 0.5\right). \tag{31}$$

The $\mathbb{A}_2(3)$ model has the following canonical representation under the physical measure $\mathcal{P}$

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ -X_{3t} \end{pmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_{1t}+\beta_{32}X_{2t}} \end{bmatrix} dW_t. \tag{32}$$

Admissibility conditions require that $\kappa_{11}\theta_1 + \kappa_{12}\theta_2 \geq 0$, $\kappa_{21}\theta_1 + \kappa_{22}\theta_2 \geq 0$, $\theta_1 \geq 0$, $\theta_2 \geq 0$, $\delta_{1(1)} \geq 0$, $\delta_{1(2)} \geq 0$, $\kappa_{12} \leq 0$, $\kappa_{21} \leq 0$, $\beta_{31} \geq 0$ and $\beta_{32} \geq 0$. The Feller condition requires that $\kappa_{11}\theta_1 + \kappa_{12}\theta_2 \geq \frac{1}{2}$ and $\kappa_{21}\theta_1 + \kappa_{22}\theta_2 \geq \frac{1}{2}$. The essentially affine $\mathbb{A}_2(3)$ specifies the market price of risk as follows

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)} X_{1t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)} X_{2t}}{\sqrt{X_{2t}}} \\ \frac{\lambda_{0(3)} + (\lambda_{0(3)}\beta_{31} + \lambda_{1(31)}) X_{1t} + (\lambda_{0(3)}\beta_{32} + \lambda_{1(32)}) X_{2t} + \lambda_{1(33)} X_{3t}}{\sqrt{1 + \beta_{31} X_{1t} + \beta_{32} X_{2t}}} \end{bmatrix}. \tag{33}$$

For the extended essentially affine $\mathbb{A}_2(3)$ model, Λ_t is specified as follows

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)} + \lambda_{1(11)} X_{1t} + \lambda_{1(12)} X_{2t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)} + \lambda_{1(21)} X_{1t} + \lambda_{1(22)} X_{2t}}{\sqrt{X_{2t}}} \\ \frac{\lambda_{0(3)} + \lambda_{1(31)} X_{1t} + \lambda_{1(32)} X_{2t} + \lambda_{1(33)} X_{3t}}{\sqrt{1 + \beta_{31} X_{1t} + \beta_{32} X_{2t}}} \end{bmatrix}.$$

As for the $\mathbb{A}_1(3)$ model, both specification are in general not nested, but the extended specification offers some extra degrees of freedom provided that the Feller condition is satisfied.

The $\mathbb{A}_3(3)$ model has the following canonical representation under the physical measure $\mathcal{P}$

$$d \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{pmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ \theta_3 - X_{3t} \end{pmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{X_{2t}} \end{bmatrix} dW_t. \tag{34}$$

Admissibility conditions require that $\kappa_{11}\theta_1 + \kappa_{12}\theta_2 + \kappa_{13}\theta_3 \geq 0$, $\kappa_{21}\theta_1 + \kappa_{22}\theta_2 + \kappa_{23}\theta_3 \geq 0$, $\kappa_{31}\theta_1 + \kappa_{32}\theta_2 + \kappa_{33}\theta_3 \geq 0$, $\theta_1 \geq 0$, $\theta_2 \geq 0$, $\theta_3 \geq 0$, $\delta_{1(1)} \geq 0$, $\delta_{1(2)} \geq 0$, $\delta_{1(3)} \geq 0$ and $\kappa_{ij} \leq 0$, $\forall i \neq j$. The Feller condition requires that $\kappa_{11}\theta_1 + \kappa_{12}\theta_2 + \kappa_{13}\theta_3 \geq \frac{1}{2}$, $\kappa_{21}\theta_1 + \kappa_{22}\theta_2 + \kappa_{23}\theta_3 \geq \frac{1}{2}$ and $\kappa_{31}\theta_1 + \kappa_{32}\theta_2 + \kappa_{33}\theta_3 \geq \frac{1}{2}$. Unlike the $\mathbb{A}_1(3)$ and $\mathbb{A}_2(3)$ models, the essentially and the extended essentially affine $\mathbb{A}_3(3)$ models are nested.[24] For the essentially affine model, the market price of risk is specified as follows

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)} X_{1t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)} X_{2t}}{\sqrt{X_{2t}}} \\ \frac{\lambda_{0(3)} X_{3t}}{\sqrt{X_{3t}}} \end{bmatrix}. \tag{35}$$

The extended essentially affine model specifies Λ_t as follows

$$\Lambda_t = \begin{bmatrix} \frac{\lambda_{0(1)} + \lambda_{1(11)} X_{1t} + \lambda_{1(12)} X_{2t} + \lambda_{1(13)} X_{3t}}{\sqrt{X_{1t}}} \\ \frac{\lambda_{0(2)} + \lambda_{1(21)} X_{1t} + \lambda_{1(22)} X_{2t} + \lambda_{1(23)} X_{3t}}{\sqrt{X_{2t}}} \\ \frac{\lambda_{0(3)} + \lambda_{1(31)} X_{1t} + \lambda_{1(32)} X_{2t} + \lambda_{1(33)} X_{3t}}{\sqrt{X_{3t}}} \end{bmatrix}. \tag{36}$$

[24] Note that the completely affine $\mathbb{A}_3(3)$ coincides with its essentially counterpart.

To estimate the extended affine $\mathbb{A}_3(3)$ model, we use the following alternative canonical representation

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} = \begin{pmatrix} \kappa\theta_1 \\ \kappa\theta_2 \\ \kappa\theta_3 \end{pmatrix} - \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{X_{2t}} \end{bmatrix} dW_t.$$

Using this equivalent representation, we make a change of variable that guarantees that the Feller condition is not violated

$$\kappa\theta_i = \left(\kappa\theta_i^{*2} + 0.5\right)$$

B Conditional Moments of the State Vector

We compute explicit expressions for the two first conditional moments following Fackler (2000) who extends the formula provided by Fisher and Gilles (1996).

B.1 Conditional Expectation

The integral form of the stochastic differential equation (1) under the actual probability measure $\mathcal{P}$ is

$$X_{t+\tau} = X_t + \int_t^{t+\tau} \kappa\left(\theta - X_u\right) du + \int_t^{t+\tau} \Sigma\sqrt{S_u} dW_u. \tag{37}$$

Applying the Fubini theorem, we get

$$E_t\left[X_{t+\tau}\right] = X_t + \int_t^{t+\tau} \kappa\left(\theta - E_t\left[X_u\right]\right) du.$$

Differentiating with respect to τ implies the following ODE

$$\frac{dE_t\left[X_{t+\tau}\right]}{d\tau} = \kappa\theta - \kappa E_t\left[X_{t+\tau}\right], \tag{38}$$

with the initial condition

$$E_t\left[X_t\right] = X_t. \tag{39}$$

The solution to this ODE has the following form

$$E_t\left[X_{t+\tau}\right] = a\left(t,\tau\right) + b\left(t,\tau\right) X_t. \tag{40}$$

Making the identification with (37) yields the following ODE's

$$\frac{\partial a\left(t,\tau\right)}{\partial\tau} = \kappa\theta - \kappa a\left(t,\tau\right) \tag{41}$$

and

$$\frac{\partial b\left(t,\tau\right)}{\partial\tau} = -\kappa b\left(t,\tau\right), \tag{42}$$

with the initial conditions

$$a(t,\tau) = 0 \text{ and } b(t,\tau) = I_N.$$

If the matrix κ is *non-singular*, the solution of equations (41) and (42) are

$$a(t,\tau) = (I_N - \exp(-\kappa\tau))\,\theta \text{ and } b(t,\tau) = \exp(-\kappa\tau), \tag{43}$$

where $\exp(-\kappa(\tau - t))$ is given by the power series

$$\exp(-\kappa\tau) = I - \tau\kappa + \frac{\tau^2}{2!}\kappa^2 + \cdots.$$

Combining these expressions with (40), we get

$$E_t[X_{t+\tau}] = (I_N - \exp(-\kappa\tau))\,\theta + \exp(-\kappa\tau)\,X_t. \tag{44}$$

One can notice that if the eigenvalues of the matrix κ are strictly positive, then

$$\lim_{\tau\to\infty} \exp(-\kappa\tau) = 0,$$

and the unconditional expectation of $X_{t+\tau}$ is given by

$$E[X_t] = \theta, \forall t.$$

B.2 Conditional Variance

Applying Itô's lemma to (44) yields

$$dE_t[X_{t+\tau}] = b(t,\tau)\,\Sigma\sqrt{S_t}dW_t,$$

or equivalently

$$X_{t+\tau} = E_t[X_{t+\tau}] + \int_t^{t+\tau} b(u, t+\tau-u)\,\Sigma\sqrt{S_u}dW_u.$$

Under some technical conditions[25]

$$\begin{aligned} var_t[X_{t+\tau}] &= var_t\left[\int_t^{t+\tau} b(u, t+\tau-u)\,\Sigma\sqrt{S_u}dW_u\right] \\ &= E_t\left[\int_t^{t+\tau} b(u, t+\tau-u)\,\Sigma S_u \Sigma b(u, t+\tau-u)^\top du\right] \\ &= \int_t^{t+\tau} b(u, t+\tau-u)\,\Sigma diag(\alpha + \mathcal{B}E_t[X_u])\,\Sigma b(u, t+\tau-u)^\top du. \end{aligned} \tag{45}$$

Following Fackler (2000), the vectorized version of (45) is

$$vec(var_t[X_{t+\tau}]) = \int_t^{t+\tau} (b(u, t+\tau-u) \otimes b(u, t+\tau-u))\,(\Sigma\otimes\Sigma)\,\mathcal{D}(\alpha + \mathcal{B}E_t[X_u])\,du, \tag{46}$$

[25] See Neftci (1996).

where $\otimes$ denotes the Kronecker product operator and $\mathcal{D}$ is a $n^2 \times n$ matrix such that

$$\mathcal{D}_{ij} = \begin{cases} 1 \text{ if } i = (j-1)\,n + j \\ 0 \qquad\qquad \text{otherwise} \end{cases}. \tag{47}$$

In the case of a 3 factor model, $\mathcal{D}$ is expressed as

$$\mathcal{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{48}$$

Using (44), expression (46) can be rearranged as follows

$$vec\,(var_t\,[X_{t+\tau}]) = v_0(t,\tau) + v_1(t,\tau)X_t,$$

where

$$v_0(t,\tau) = \int_t^{t+\tau} (b\,(u, t+\tau-u) \otimes b\,(u, t+\tau-u))\,(\Sigma \otimes \Sigma)\,\mathcal{D}\,(\alpha + \mathcal{B}a\,(t, u-t))\,du \tag{49}$$

and

$$v_1(t,\tau) = \int_t^{t+\tau} (b\,(u, t+\tau-u) \otimes b\,(u, t+\tau-u))\,(\Sigma \otimes \Sigma)\,\mathcal{D}\mathcal{B}b\,(t, u-t)\,du. \tag{50}$$

Differentiating (49) and (50) with respect to τ yields the following ODE's

$$\frac{\partial v_0(t,\tau)}{\partial \tau} = (\Sigma \otimes \Sigma)\,\mathcal{D}\,(\alpha + \mathcal{B}a\,(t,\tau)) - (\kappa \otimes I_N + I_N \otimes \kappa)\,v_0(t,\tau), \tag{51}$$

and

$$\frac{\partial v_1(t,\tau)}{\partial \tau} = (\Sigma \otimes \Sigma)\,\mathcal{D}\mathcal{B}b\,(t,\tau) - (\kappa \otimes I_N + I_N \otimes \kappa)\,v_1(t,\tau) \tag{52}$$

Combining these ODEs with equations (41) and (42), we get the following two systems of ODE's

$$\begin{bmatrix} \frac{\partial a(t,\tau)}{\partial \tau} \\ \frac{\partial v_0(t,\tau)}{\partial \tau} \end{bmatrix} = \Theta - \boldsymbol{\kappa} \begin{bmatrix} a\,(t,\tau) \\ v_0(t,\tau) \end{bmatrix}, \tag{53}$$

and

$$\begin{bmatrix} \frac{\partial b(t,\tau)}{\partial \tau} \\ \frac{\partial v_1(t,\tau)}{\partial \tau} \end{bmatrix} = -\boldsymbol{\kappa} \begin{bmatrix} b\,(t,\tau) \\ v_1(t,\tau) \end{bmatrix}, \tag{54}$$

where

$$\Theta = \begin{bmatrix} \kappa\theta \\ (\Sigma \otimes \Sigma)\,\mathcal{D}\alpha \end{bmatrix} \tag{55}$$

and

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa & 0 \\ -(\Sigma \otimes \Sigma)\,\mathcal{DB} & (\kappa \otimes I_N + I_N \otimes \kappa) \end{bmatrix}. \tag{56}$$

The initial conditions are $a\,(t,\tau) = 0$, $b\,(t,\tau) = I_N, v_0(t,0) = 0$ and $v_1(t,0)$. Provided that κ is nonsingular, the solution to these two systems is given by

$$\begin{bmatrix} a\,(t,\tau) \\ v_0(t,\tau) \end{bmatrix} = (I_N - \exp\left(-\boldsymbol{\kappa\tau}\right))\,\boldsymbol{\kappa}^{-1}\Theta, \tag{57}$$

and

$$\begin{bmatrix} b\,(t,\tau) \\ v_1(t,\tau) \end{bmatrix} = \exp\left(-\boldsymbol{\kappa\tau}\right) \begin{bmatrix} I_N \\ 0 \end{bmatrix}, \tag{58}$$

where $\exp\left(-\kappa\,(\tau - t)\right)$ is given by the power series

$$\exp\left(-\kappa\tau\right) = I - \tau\kappa + \frac{\tau^2}{2!}\kappa^2 + \cdots. \tag{59}$$

Since $\boldsymbol{\kappa}^{-1}$ can be written as

$$\begin{bmatrix} \kappa^{-1} & 0 \\ (\kappa \otimes I_N + I_N \otimes \kappa)^{-1}\,(\Sigma \otimes \Sigma)\,\mathcal{DB}\kappa^{-1} & (\kappa \otimes I_N + I_N \otimes \kappa)^{-1} \end{bmatrix}, \tag{60}$$

if we assume that the eigenvalues of $\boldsymbol{\kappa}$ are strictly positive, then

$$\lim_{\tau \to \infty} \exp\left(-\boldsymbol{\kappa\tau}\right) = 0$$

and the unconditional vectorized variance is

$$\begin{aligned} vec\,(var\,[X_t]) &= \lim_{\tau \to \infty} v_0(t,\tau) \\ &= (\kappa \otimes I_N + I_N \otimes \kappa)^{-1}\,(\Sigma \otimes \Sigma)\,\mathcal{D}\,(\mathcal{B}\theta + \alpha)\,. \end{aligned} \tag{61}$$

Computing the first two conditional moments involves the evaluation of the power series (59). Several methods for evaluating the exponential of a matrix are provided in the literature[26]. However, as pointed out by Fackler (2000), the eigenvalues decomposition, suggested by Fisher and Gilles (1996) and used by Duffee (2002), and the Padé approximation yield good results in this particular context. In this paper, we use the Padé approximation to compute the conditional expectation and variance.

C The Extended Kalman Filter

C.1 The Algorithm

The Extended Kalman Filter relies on the same state-space representation the QML method but assumes that all yields are measured with errors, which does not allow the state vector to be inverted

[26] See Moler and Van Loan (1978).

from the cross section of yields. When all yields are measured with errors, the measurement equation is rewritten as follows

$$y_t = \overline{A} + \overline{B} X_t + u_t \tag{62}$$

where y_t is a vector that contains M observed yields, $u_{t+1} \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$, Σ is a diagonal matrix with elements σ_i^2,

$$\overline{A} = \begin{bmatrix} -\frac{A(\tau_1)}{\tau_1} \\ \cdot \\ \cdot \\ -\frac{A(\tau_M)}{\tau_M} \end{bmatrix} \text{ and } \overline{B} = \begin{bmatrix} \frac{B(\tau_1)}{\tau_1} \\ \cdot \\ \cdot \\ \frac{B(\tau_M)}{\tau_M} \end{bmatrix}.$$

The transition equation is discretized as follows

$$X_{t+1} = E_t\left[X_{t+1}\right] + v\left(X_t\right)^{\frac{1}{2}} \varepsilon_{t+1}, \tag{63}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, I)$, and $v\left(X_t\right)^{\frac{1}{2}} = var_t\left[X_{t+1}\right]^{\frac{1}{2}}$ is the Cholesky decomposition of the conditional covariance matrix of the state vector.

Since the conditional expectation of the state vector is an affine function, the transition equation (63) can be rewritten as follows

$$X_{t+\Delta t} = a + bX_t + v\left(X_t\right)^{\frac{1}{2}} \varepsilon_{t+\Delta t}, \tag{64}$$

where a and b are given by (43).

Let us denote the contemporaneous forecast (or the filtered value) of the state vector and its corresponding covariance matrix by $X_{t|t}$ and $P_{t|t}$, the extended Kalman filter algorithm works as follows at any time t:[27]

1. Given $X_{t|t}$ and $P_{t|t}$, compute the one period ahead forecast of the state vector and its corresponding covariance matrix[28]

$$X_{t+1|t} = a + bX_{t|t} \tag{65}$$

and

$$P_{t+1|t} = b' P_{t|t} b + v\left(X_{t|t}\right). \tag{66}$$

2. Compute the one period ahead forecast of y_{t+1} and its corresponding covariance matrix

$$y_{t+1|t} = \overline{A} + \overline{B} X_{t+1|t} \tag{67}$$

and

$$V_{t+1|t} = \overline{B}' v\left(X_t\right) \overline{B} + \Sigma. \tag{68}$$

[27] See Hamilton (1994) for further details.

[28] We make the normalization that $\Delta t = 1$.

3.Compute the forecast error of y_{t+1}, $e_{t+1|t} = y_{t+1} - y_{t+1|t}$.

4. Update the contemporaneous forecast of the state vector and its corresponding covariance matrix

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t}\overline{B}V_{t+1|t}^{-1}e_{t+1|t} \tag{69}$$

and

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}\overline{B}V_{t+1|t}^{-1}\overline{B}'P_{t+1|t}. \tag{70}$$

5. Return to step 1.

The log quasi-likelihood of observation $t+1$ is then

$$\mathcal{L}_t(\Theta) = -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log\left(\det\left(V_{t+1|t}\right)\right) - \frac{1}{2}e'_{t+1|t}V_{t+1|t}e_{t+1|t}. \tag{71}$$

In order to start the recursion, we need the initial one period ahead forecast $X_{1|0}$ and its covariance matrix $P_{1|0}$. The unconditional two first moments are used in the first step of the recursion, which implies that

$$X_{1|0} = E[X_t] \text{ and } P_{1|0} = var[X_t].$$

C.2 Variance Forecasts

Rather than using the implied value of the state vector as in (22), the model-implied conditional variance of yields is computed using the filtered state vector $X_{t|t}$

$$var_t(y_{t+1}(\tau)) = \overline{B}(\tau)'P_{t+1|t}\overline{B}(\tau) + var(u_{t+1}(\tau)), \tag{72}$$

where

$$P_{t+1|t} = b'P_{t|t}b + var_t[X_{t+1}]. \tag{73}$$

References

[1] Almeida, C., Graveline, J. and S. Joslin, 2006, "Do Options Contain Information About Excess Bond Returns?", Working Paper, Stanford University.

[2] Andersen, T.G., and L. Benzoni, 2005, "Can Bonds Hedge Volatility in the US Treasury Market? A Specification Test for Affine Models," Working Paper, Northwestern University.

[3] Andersen, T., Bollerslev, T., Diebold, F.X. and H. Ebens, 2001, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43-67.

[4] Andersen, T., Bollerslev, T., Diebold, F.X. and P. Labys, 2001, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.

[5] Ang, A., and M. Piazzesi, 2003, "A No Arbitrage Vector Autoregression of The Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745-787.

[6] Balduzzi, P., S. Das, S., Foresi, and R. K. Sundaram, 1996, "A Simple Approach to Three-Factor Affine Models of the Term Structure," *Journal of Fixed Income*, 6, 43-53.

[7] Bikbov, R., and M. Chernov, 2004, "Term Structure and Volatility: Lessons from the Eurodollar Markets," Working Paper, Columbia Business School.

[8] Bliss, R, 1997, "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, 9, 197-231.

[9] Chen, R.R., and L. Scott, 1993, "Maximum Likelihood Estimation for Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, December, 14-31.

[10] Cheridito, P., D. Filipović, and R. Kimmel, 2005, "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, forthcoming.

[11] Cochrane, J., and M. Piazzesi, 2005, "Bond Risk Premia," *American Economic Review*, 95, 138-160.

[12] Collin-Dufresne, P., R.S. Goldstein, 2002, "Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility," *Journal of Finance*, 57, 1685-1730.

[13] Collin-Dufresne, P., R.S. Goldstein, and C.S. Jones, 2004, "Can Interest Rate Volatility be Extracted from the Cross Section of Bond yields? An Investigation of Unspanned Stochastic Volatility," NBER Working Paper 10756.

[14] Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

[15] Dai, Q., and K. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943-1978.

[16] Dai, Q., and K. Singleton, 2002, "Expectation Puzzles, Time Varying Risk Premia and Affine Models of the Term Structure," *Journal of Financial Economics*, 63, 415-441.

[17] Dai, Q., and K. Singleton, 2003, "Term Structure Modeling in Theory and Reality," *Review of Financial Studies*, 16, 631-678.

[18] Dai, Q., K. Singleton, and W. Yang, 2003, "Is Regime Shifting Priced in the US Treasury Market?," Working Paper, New York University.

[19] Duan J. C. , and J. G. Simonato, 1999, "Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter," *Review of Quantitative Finance and Accounting*, 13, 111–135

[20] Duarte, J., 2004, "Evaluating an Alternative Risk Preference in Affine Term Structure Models," *Review of Financial Studies,* 17, 379-404.

[21] Duffee, G., 1996, "Idiosyncratic Variation of Treasury Bill Yields," *Journal of Finance*, 51, 527-552.

[22] Duffee, G., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405-443.

[23] Duffee, G., 2005, "Term Structure Estimation Without Using Latent Factors," *Journal of Financial Economics*, forthcoming.

[24] Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379-406.

[25] Fackler, P., 2000, "Moments of Affine Diffusions," Working Paper, North Carolina State University.

[26] Fama, E.E, and R.R., Bliss, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

[27] Fan, R., A. Gupta, and P. Ritchken, 2003, "Hedging in the Possible Presence of Unspanned Stochastic Volatility: Evidence from Swaption Markets," *Journal of Finance* 58, 2219-2248.

[28] Fisher, M., and C., Gilles, 1996, "Estimating Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Board.

[29] Gallant, A.R., and G. Tauchen, 1998, "Reprojecting Partially Observed Systems with Applications to Interest Rate Diffusions," *Journal of American Statistical Association*, 93, 10-24.

[30] Jaganathan, R., A. Kaplin, and S.G. Sun, 2003, "An Evaluation of Multi-Factor CIR Models Using LIBOR, Swap Rates, Cap and Swaption Prices," *Journal of Econometrics*, 116, 113-146.

[31] Hamilton, J., 1994, "*Time Series Analysis*," Princeton University Press.

[32] Litterman, R., and J. Scheinkman, 1991. "Common Factors Affecting Bond returns," *Journal of Fixed Income* 1, 54-61.

[33] Li, H., and F. Zhao, 2003, "Unspanned Stochastic Volatility, Is It There After All? Evidence From Hedging Interest Rate Caps," *Journal of Finance*, forthcoming.

[34] Liu, J., F. Longstaff and R. Mandell, 2006, "The Market Price of Credit Risk: An Empirical Analysis of Interest Rate Swap Spreads," *Journal of Business*, forthcoming.

[35] McCulloch, J. H., 1975, "The Tax-Adjusted Yield Curve," *Journal of Finance*, 30, 811-830.

[36] McCulloch, J. H., and H. C. Kwon, 1991, "US Term Structure Data, 1947-1991," Working Paper, Ohio State University.

[37] Modigliani, F., and R. Sutch, 1967, "Debt Management and the Term Structure of Interest Rates: An Empirical Analysis of Recent Experience," *Journal of Political Economy*, 75, 569-589.

[38] Moler, C., and C. F.,Van Loan, 1978, "Nineteen Dubious Ways to Compute the Exponential of a Matrix," *SIAM Review*, 20, 801-836.

[39] Neftci, S., 1996, *An Introduction to the Mathematics of Financial Derivatives*, Academic Press.

[40] Newey, W., and K. West, 1987, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

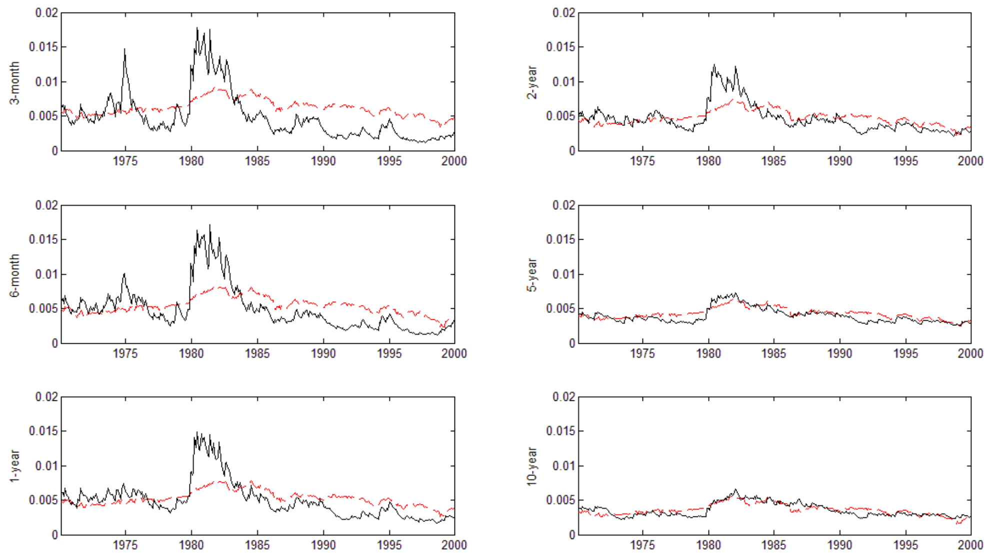
[41] Piazzesi, M., 2001, "Macroeconomic Jump Effects and the Yield Curve," Working paper, UCLA.

[42] Piazzesi, M., 2005, "Affine Term Structure Models," forthcoming in Y. Aït Sahalia, and L.P. Hansen, ed.: *Handbook of Financial Econometrics,* Elsevier Science, Amsterdam.

[43] Schwert, G.W., 1989, "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44, 1115-1153.

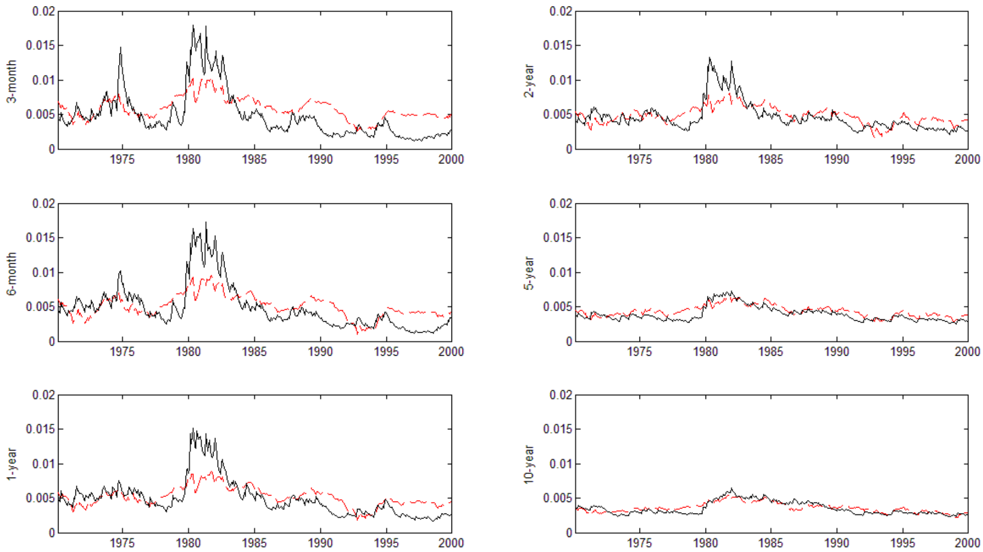
[44] Thompson, S., 2004, "Identifying Term Structure Volatility from the LIBOR-Swap Curve", Working Paper, Harvard University.

Figure 1: Conditional Volatility Implied by the Essentially Affine $\mathbb{A}_1(3)$ Model and EGARCH(1,1) Estimates for Various Maturities.



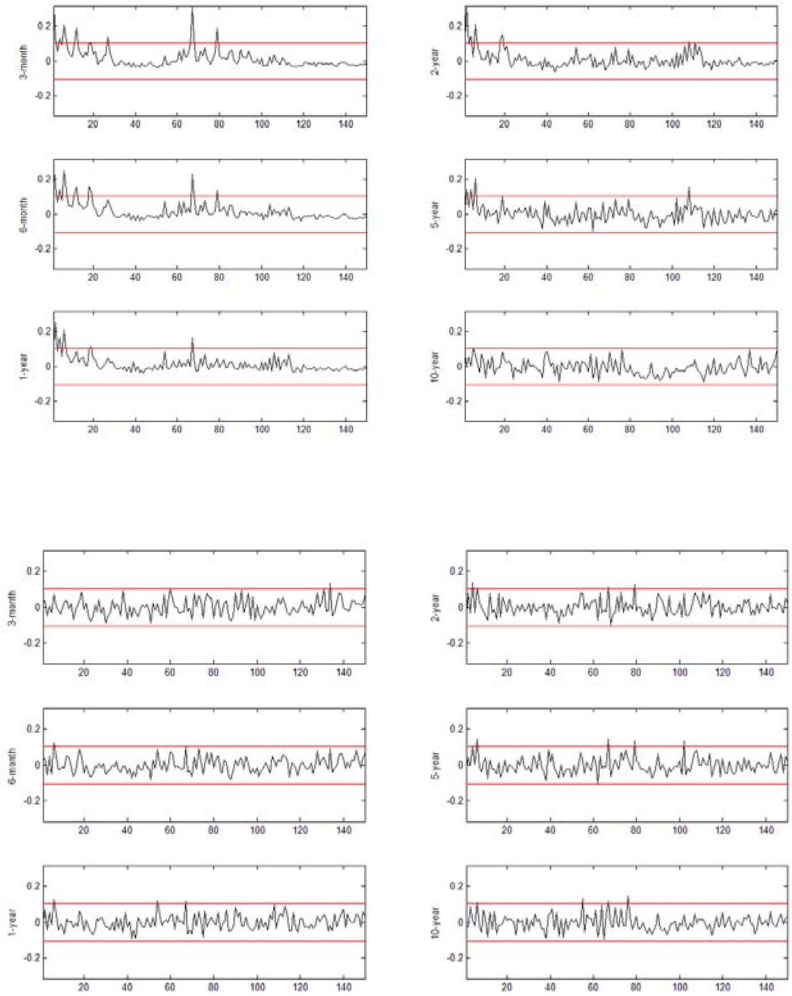
Notes to figure: For each maturity, the dashed line depicts the conditional volatility implied by the essentially affine $\mathbb{A}_1(3)$ model and the solid line depicts the EGARCH (1,1) volatility estimates of changes in yields. The EGARCH(1,1) is estimated assuming that the conditional mean of changes in yields is generated by an AR(1) process. The sample period is from January 1970 to December 1999.

Figure 2: Conditional Volatility Implied by the Extended Essentially Affine $\mathbb{A}_3(3)$ Model and EGARCH(1,1) Estimates for Various Maturities.



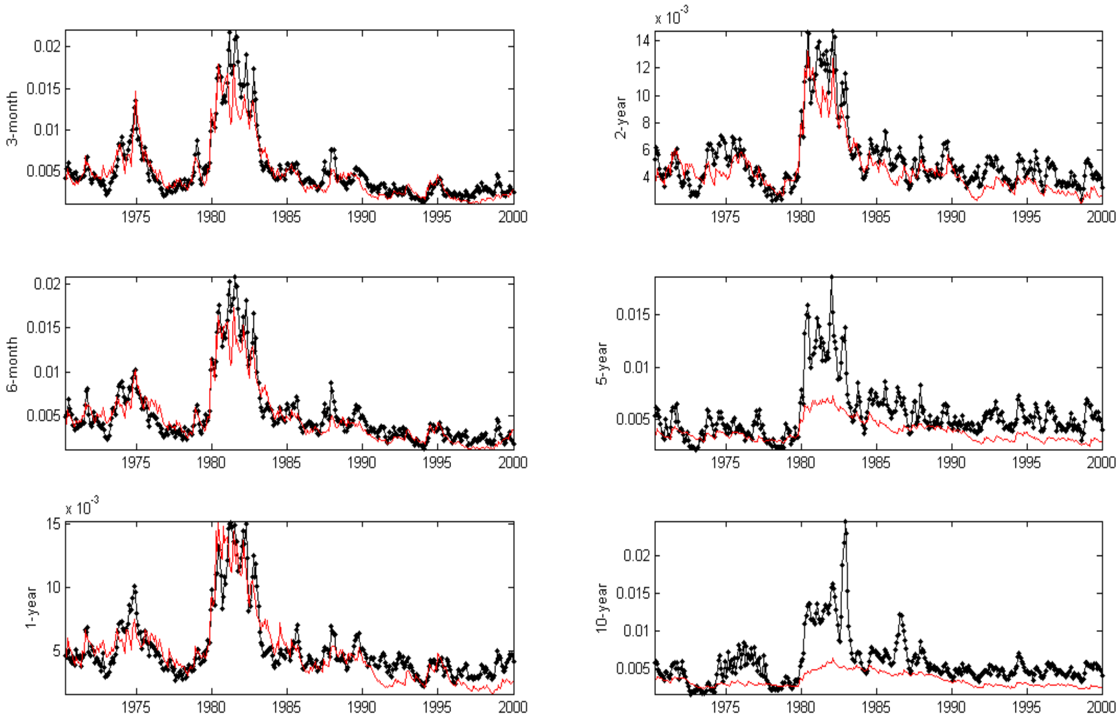
Notes to figure: For each maturity, the dashed line depicts the conditional volatility implied by the extended essentially affine $\mathbb{A}_3(3)$ model, and the solid line depicts the EGARCH (1,1) volatility estimates of changes in yields. The EGARCH(1,1) is estimated assuming that the conditional mean of changes in yields is generated by an AR(1) process. The sample period is from January 1970 to December 1999.

Figure 3: Autocorrelation of the Standardized Squared Change in Yields Implied by the Essentially Affine $\mathbb{A}_1(3)$ Model and the EGARCH Estimates.



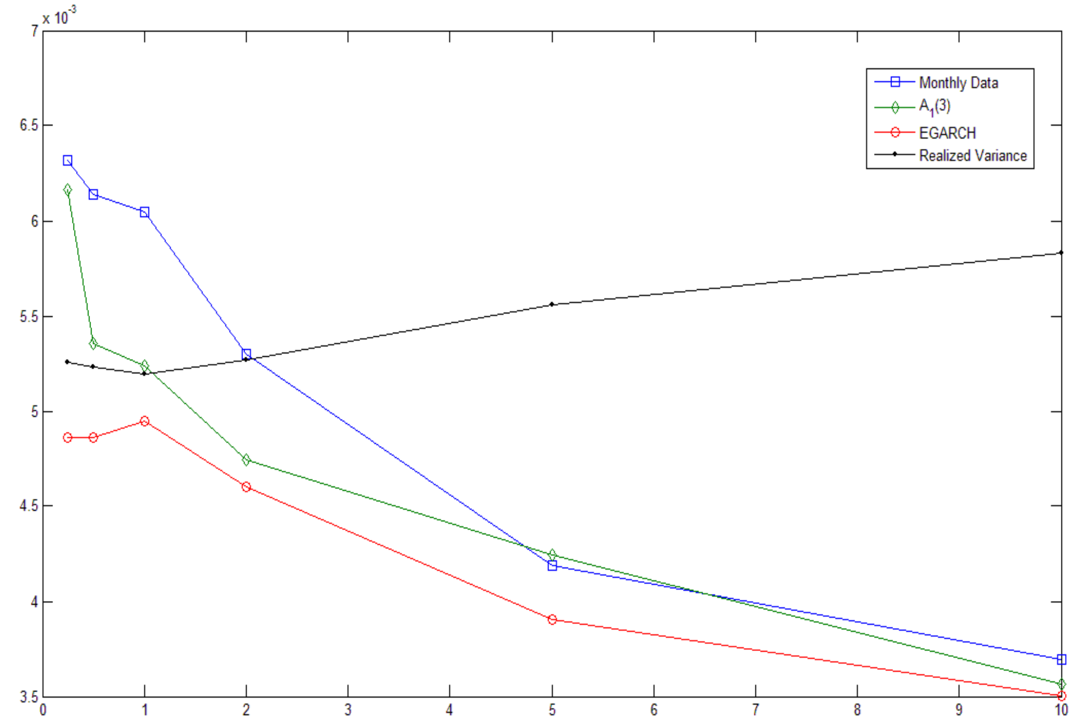
Notes to figure: For each maturity, we plot the autocorrelation path of the standardized squared changes in yields implied by the essentially affine $\mathbb{A}_1(3)$ model (top panel) and EGARCH model (bottom panel) with the Bartlett standard errors.

Figure 4: Conditional Volatility Implied by the Realized Variance Model and the EGARCH(1,1) Estimates for Various Maturities



Notes to figure: For each maturity, the solid line depicts the EGARCH (1,1) volatility estimates of changes in yields, and the diamonds depict the volatility implied by the realized variance model. The EGARCH(1,1) is estimated assuming that the conditional mean of changes in yields is generated by an AR(1) process. The sample period is from January 1970 to December 1999.

Figure 5: Term Structure of Unconditional Yield Volatility.



Notes to figure: The monthly data unconditional volatility is computed as the standard deviation of changes in monthly yields. For each model, the unconditional volatility is computed as the average of the conditional volatility paths.

Table 1: Parameter Estimates for Completely Affine Models

Parameter	Models								
	$A_1(3)$			$A_2(3)$			$A_3(3)$		
	Factor			Factor			Factor		
	1	2	3	1	2	3	1	2	3
δ_0	0.0219			0.0260			0.02525		
	(0.1273)			(0.0272)			(0.0023)		
δ_{1j}	0.0015	0.0006	0.0042	0.0013	0.0020	0.0087	0.00098	0.00001	0.01356
	(0.001)	(0.0000)	(0.0041)	(0.0024)	(0.0026)	(0.0031)	(0.0006)	(0.0007)	(0.0069)
κ_{1j}	0.0297	0	0	0.1344	-0.2814	0	0.58200	-0.14999	0
	(0.0353)			(0.0516)	(0.0319)		(0.0282)	(0.0135)	
κ_{2j}	-0.0951	0.3225	17.6585	-0.1691	0.4416	0	-0.02339	0.06580	-0.29900
	(4.8206)	(0.2352)	(11.1584)	(0.1856)	(0.2631)		(0.1369)	(0.0268)	(0.3018)
κ_{3j}	0.0496	-0.0184	1.8632	0.4313	-1.6688	1.5951	-1.08000	0	1.56000
	(0.2611)	(0.0166)	(0.2699)	(0.3362)	(0.4952)	(0.1775)	(0.0550)		(0.0840)
θ_j	5.5100	0	0	2.9584	0.8271	0	5.24000	13.70000	1.85000
	(35.0114)			(0.9364)	(0.1439)		(3.8613)	(2.7803)	(0.1457)
λ_{0j}	-0.0405	-0.0167	-0.1143	-0.0418	-0.0063	-0.2228	-0.05262	-0.00118	-0.25601
	(0.0262)	(0.0404)	(0.2522)	(0.1463)	(0.1032)	(0.2551)	(0.1992)	(0.1482)	(1.5073)
β_1	1	0	0	1	0	0	1	0	0
β_2	42.0946	0	0	0	1	0	0	1	0
	(14.7709)								
β_3	0.3204	0	0	0	1	0	0	0	1
	(0.1771)								
L_{1j}	0.0024	0	0	0.0023	0	0	0.00229	0	0
	(0.0001)			(0.0001)			(0.0001)		
L_{2j}	-0.0006	0.0013	0	-0.0006	0.0013	0	-0.00061	0.00132	0
	(0.0001)	(0.0001)		(0.0001)	(0.0001)		(0.0001)	(0.0001)	
L_{3j}	0.0004	-0.0001	0.0016	0.0004	-0.0002	0.0016	0.00030	-0.00015	0.00156
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)
Log Likelihood		10149.13			10170.30			10219.31	

Notes: We estimate the models using QML on a sample of monthly data from January 1970 to December 1999. Yields with maturities of 6 months, 2 years and 10 years are assumed to be measured exactly, whereas yields with maturities of 3 months, 1 year and 5 years are assumed to be measured with errors. Zero-coupon yields are interpolated using the unsmoothed Fama and Bliss method. For the $A_2(3)$ model, β_{31} is set equal to zero and β_{32} is set equal to one. For the $A_3(3)$ model, we constrain the coefficients κ_{13} and κ_{32} to be equal to zero. Standard errors are given in parentheses.

Table 2: Parameter Estimates for Essentially Affine Models

Parameter	Model					
	$A_1(3)$			$A_2(3)$		
	Factor			Factor		
	1	2	3	1	2	3
δ_0	0.0363 (0.0310)			0.0260 (0.0434)		
δ_{1j}	0.0023 (0.0018)	0.0018 (0.0011)	0.0033 (0.0007)	0.0013 (0.0039)	0.0020 (0.0036)	0.0087 (0.0025)
κ_{1j}	0.0338 (0.0067)	0	0	0.1340 (0.0111)	-0.2810 (0.0219)	0
κ_{2j}	-0.0504 (0.0202)	0.4075 (0.1788)	2.8481 (0.1121)	-0.1690 (0.0317)	0.4420 (0.1234)	0
κ_{3j}	0.2295 (0.0186)	-0.0287 (0.0366)	2.9503 (0.0410)	0.4310 (0.0968)	-1.6700 (0.1344)	1.6000 (0.1390)
θ_j	5.2514 (7.4337)	0	0	2.9600 (0.8815)	0.8270 (0.2765)	0.0000
λ_{0j}	-0.0488 (0.1233)	-6.0024 (9.1750)	0.2481 (2.0312)	-0.0418 (0.3251)	-0.0063 (0.0829)	-0.2230 (0.3458)
$\lambda_{1(1j)}$	0	0	0	0	0	0
$\lambda_{1(2j)}$	62.5415 (7.3907)	0.0922 (0.1890)	5.7439 (1.2761)	0	0	0
$\lambda_{1(3j)}$	-0.1963 (0.5731)	0.0127 (0.7569)	-1.8131 (0.0686)	-3.8E-08 (0.0538)	-9.3E-09 (0.9544)	1.2E-08 (1.5732)
β_1	1.0000	0	0	1	0	0
β_2	10.3841 (5.8989)	0	0	0	1	0
β_3	0.2859 (0.9838)	0	0	0	1	0
L_{1j}	0.0023 (0.0002)	0	0	0.0023 (0.0001)	0	0
L_{2j}	-0.0006 (0.0003)	0.0013 (0.0001)	0	-0.0006 (0.0002)	0.0013 (0.0001)	0
L_{3j}	0.0001 (0.0013)	-0.0001 (0.0002)	0.0016 (0.0003)	0.0004 (0.0007)	-0.0002 (0.0002)	0.0016 (0.0001)
Log Likelihood	10179.00			10170.23		

Notes: We estimate the models using QML on a sample of monthly data from January 1970 to December 1999. Yields with maturities of 6 months, 2 years and 10 years are assumed to be measured exactly, whereas yields with maturities of 3 months, 1 year and 5 years are measured with errors. Zero-coupon yields are interpolated using the unsmoothed Fama and Bliss method. For the $A_2(3)$ model, β_{31} is set equal to zero and β_{32} is set equal to one. Standard errors are given in parentheses.

Table 3: Parameter Estimates for Extended Essentially Affine Models

Parameter	Model					
	$A_1(3)$			$A_3(3)$		
	Factor			Factor		
	1	2	3	1	2	3
δ_0	0.0321			0.0244		
	(0.0178)			(0.0008)		
δ_{1j}	0.0021	0.0017	0.0052	0.0007	0.0002	0.0126
	(0.0016)	(0.0010)	(0.0052)	(0.0005)	(0.0000)	(0.0007)
κ_{1j}	0.0667	0	0	1.6564	-0.2052	-0.1434
	(0.1256)			(0.0392)	(0.0151)	(0.0950)
κ_{2j}	-0.2809	0.5031	4.7426	-0.0066	0.3561	-1.3930
	(0.2090)	(0.2310)	(1.0343)	(0.0518)	(0.0809)	(0.0865)
κ_{3j}	0.0620	-0.0543	3.1463	-1.3150	0	1.1875
	(0.1533)	(0.1481)	(0.4839)	(0.0513)		(0.1250)
$\kappa\theta^*_j$	0.1576	0	0	0.0064	0.0271	0.0033
	(0.3634)			(0.3705)	(0.1989)	(0.1098)
λ_{0j}	0.3346	-6.0298	0.0104	0.2093	0.4104	1.0452
	(0.9051)	(23.4110)	(1.8343)	(0.0345)	(0.2528)	(0.0797)
$\lambda_{1(1j)}$	-0.0833	0	0	-0.9942	0.1253	0.0857
	(0.2825)			(0.0449)	(0.0119)	(0.2409)
$\lambda_{1(2j)}$	0.0056	-0.1407	5.8465	0.2578	-0.3271	1.0821
	(1.4697)	(1.4652)	(1.6992)	(0.2335)	(0.1983)	(0.1769)
$\lambda_{1(3j)}$	-0.0245	0.0286	-1.8283	-0.7061	0	-0.1196
	(0.3413)	(0.3319)	(1.6672)	(0.3631)		(0.1469)
β_1	1	0	0	1	0	0
β_2	10.5514	0	0	0	1	0
	(9.0713)					
β_3	0.1576	0	0	0	0	1
	(0.3230)					
L_{1j}	0.00229	0	0	0.00237	0	0
	(0.0001)			(0.0001)		
L_{2j}	-0.0006	0.0013	0	-0.0006	0.0013	0
	(0.0001)	(0.0001)		(0.0001)	(0.0001)	
L_{3j}	-0.0001	0.0001	0.0016	0.0002	-0.0002	0.0016
	(0.0007)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Log Likelihood	10174.75			10219.05		

Notes: We estimate the models using QML on a sample of monthly data from January 1970 to December 1999. Yields with maturities of 6 months, 2 years and 10 years are assumed to be measured exactly, whereas yields with maturities of 3 months, 1 year and 5 years are measured with errors. Zero-coupon yields are interpolated using the unsmoothed Fama and Bliss method. For the $A_3(3)$ model, we constrain κ_{32} to be equal to zero. Standard errors are given in parentheses.

Table 4: Regressions of EGARCH Volatility Estimates of Changes in Yields on the Conditional Volatility Implied by Different Models

Panel A: Correlations for Different Models

	Maturity					
Model	**3-month**	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
$CA_1(3)$	0.5863	0.6283	0.6651	0.667	0.8199	0.8217
$CA_2(3)$	0.6472	0.6678	0.7025	0.6593	0.8265	0.8246
$CA_3(3)$	0.7123	0.7077	0.7466	0.6705	0.8303	0.8285
$EA_1(3)$	0.5935	0.6334	0.6733	0.6769	0.8147	0.8145
$EA_2(3)$	0.6475	0.6685	0.7033	0.6605	0.8259	0.8241
$ExA_1(3)$	0.5895	0.6278	0.6665	0.6654	0.8175	0.8117
$ExA_3(3)$	0.7062	0.6949	0.7344	0.6544	0.8168	0.8228

Panel B: Regression Results for the Essentially Affine $A_1(3)$ Model

	3-month	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
Intercept	-0.0059 (-2.9950)	-0.0043 (-2.7310)	-0.0039 (-2.8555)	-0.0018 (-1.7896)	-0.0008 (-1.7816)	0.0001 (-0.3263)
Slope	1.7460 (5.0578)	1.7012 (5.3357)	1.6852 (5.9075)	1.3449 (5.7492)	1.1031 (10.0554)	0.9577 (13.0142)
Correlation	0.5935	0.6334	0.6733	0.6769	0.8147	0.8145

Panel C: Regression Results for the Extended Affine $A_3(3)$ Model

	3-month	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
Intercept	-0.0041 (-3.7889)	-0.0022 (-2.2509)	-0.0028 (-2.7312)	-0.0008 (-0.8754)	-0.0007 (-1.7173)	-0.0005 (-1.9061)
Slope	1.5261 (7.6416)	1.3991 (6.6697)	1.4897 (7.1863)	1.1059 (5.7024)	1.0639 (10.5341)	1.1264 (15.3460)
Correlation	0.7062	0.6949	0.7344	0.6544	0.8168	0.8228

Notes: We regress conditional volatility implied by EGARCH estimates of changes in yields on the conditional volatility implied by different models. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.

Table 5: Regressions of EGARCH Volatility Estimates of Yields on the Conditional Volatility Implied by Different Models

Panel A: Correlations for Different Models

	Maturity					
Model	**3-month**	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
$CA_1(3)$	0.6265	0.6543	0.6819	0.7410	0.8494	0.9207
$CA_2(3)$	0.6883	0.7024	0.7340	0.7624	0.8855	0.9300
$CA_3(3)$	0.7655	0.7521	0.7827	0.7753	0.8897	0.9330
$EA_1(3)$	0.6297	0.6541	0.6836	0.7384	0.8477	0.9135
$EA_2(3)$	0.6887	0.7031	0.7347	0.7634	0.8851	0.9296
$ExA_1(3)$	0.6305	0.6548	0.6843	0.7396	0.8489	0.9145
$ExA_3(3)$	0.7586	0.7391	0.7707	0.7591	0.8791	0.9287

Panel B: Regression Results for the Essentially Affine $A_1(3)$ Model

	3-month	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
Intercept	-0.0068 (-3.5509)	-0.0045 (-3.0735)	-0.0042 (-3.1263)	-0.0025 (-2.6432)	-0.0013 (-3.0611)	-0.0004 (-1.9148)
Slope	1.8914 (5.6270)	1.7576 (5.8031)	1.7474 (6.2239)	1.4963 (6.9458)	1.2334 (12.1511)	1.1044 (18.5498)
Correlation	0.6297	0.6541	0.6836	0.7384	0.8477	0.9135

Panel C: Regression Results for the Extended Affine $A_3(3)$ Model

	3-month	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
Intercept	-0.0049 (-4.5419)	-0.0026 (-2.6968)	-0.0032 (-3.1918)	-0.0016 (-1.8950)	-0.0014 (-3.5811)	-0.0011 (-5.3461)
Slope	1.6600 (8.5499)	1.4739 (7.3507)	1.5793 (7.7505)	1.2849 (7.1066)	1.2339 (13.4746)	1.3015 (22.8995)
Correlation	0.7586	0.7391	0.7707	0.7591	0.8791	0.9287

Notes: We regress conditional volatility implied by EGARCH estimates of yields on the conditional volatility implied by different models. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.

Table 6: Correlation Matrices for Conditional Volatility of Changes in Yields

Panel A. EGARCH Volatility Estimates

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.0000	0.9828	0.9553	0.8862	0.7700	0.5713
6-month		1.0000	0.9808	0.9320	0.8147	0.6325
1-year			1.0000	0.9631	0.8600	0.6688
2-year				1.0000	0.9127	0.7404
5-year					1.0000	0.9083
10-year						1.0000

Panel B. Volatility Implied by the Completely Affine $A_2(3)$ Model

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.0000	0.9996	0.9995	0.9991	0.9962	0.9585
6-month		1.0000	1.0000	0.9999	0.9944	0.9536
1-year		1.0000	1.0000	0.9999	0.9941	0.9526
2-year				1.0000	0.9938	0.9523
5-year					1.0000	0.9798
10-year						1.0000

Panel C. Volatility Implied by the Essentially Affine $A_2(3)$ Model

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.0000	0.9996	0.9996	0.9992	0.9968	0.9632
6-month		1.0000	1.0000	0.9999	0.9954	0.9589
1-year			1.0000	0.9999	0.9951	0.9581
2-year				1.0000	0.9948	0.9577
5-year					1.0000	0.9815
10-year						1.0000

Panel D. Volatility Implied by the Completely Affine $A_3(3)$ Model

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.0000	0.9975	0.9839	0.9418	0.8816	0.7979
6-month		1.0000	0.9926	0.9601	0.9020	0.8138
1-year			1.0000	0.9866	0.9434	0.8582
2-year				1.0000	0.9771	0.8975
5-year					1.0000	0.9674
10-year						1.0000

Panel E. Volatility Implied by the Extended Essentially Affine $A_3(3)$ Model

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
3-month	1.0000	0.9961	0.9789	0.9272	0.8662	0.7931
6-month		1.0000	0.9906	0.9511	0.8916	0.8117
1-year			1.0000	0.9840	0.9426	0.8677
2-year				1.0000	0.9805	0.9109
5-year					1.0000	0.9698
10-year						1.0000

Notes: We report the unconditional correlation of the EGARCH estimates on changes in yields and the conditional volatility implied by the $A_2(3)$ and $A_3(3)$ models at different horizons.

Table 7: Correlation Matrix for Yields

Panel A: Correlation Matrix of Yields (Raw Data)

Maturity	3-month	1-year	5-year
3-month	1	0.9845	0.8976
1-year		1	0.9424
5-year			1

Panel B: Correlation Matrix of Yields Implied by the Essentially Affine $A_1(3)$ Model

Maturity	3-month	1-year	5-year
3-month	1	0.9872	0.895
1-year		1	0.948
5-year			1

Panel C: Correlation Matrix of Yields Implied by the Extended Essentially Affine $A_3(3)$ Model

Maturity	3-month	1-year	5-year
3-month	1	0.9861	0.8964
1-year		1	0.951
5-year			1

Notes: We report the unconditional correlation of the data yields and the yields implied by the essentially affine $A_1(3)$ and the extended essentially affine $A_3(3)$.

Table 8: Correlations between Conditional Volatilities of Changes in Yields and Factors Affecting the Yield Curve

Panel A: Correlation with the Level Factor

Maturity	3-month		6-month		1-year		2-year		5-year		10-year	
EGARCH	0.729	(0.000)	0.742	(0.000)	0.774	(0.000)	0.714	(0.000)	0.796	(0.000)	0.664	(0.000)
$CA_1(3)$	0.754	(0.000)	0.751	(0.000)	0.752	(0.000)	0.751	(0.000)	0.754	(0.000)	0.751	(0.000)
$CA_2(3)$	0.887	(0.000)	0.883	(0.000)	0.884	(0.000)	0.882	(0.000)	0.877	(0.000)	0.833	(0.000)
$CA_3(3)$	0.984	(0.000)	0.979	(0.000)	0.975	(0.000)	0.941	(0.000)	0.900	(0.000)	0.832	(0.000)
$EA_1(3)$	0.758	(0.000)	0.752	(0.000)	0.755	(0.000)	0.754	(0.000)	0.759	(0.000)	0.752	(0.000)
$EA_2(3)$	0.886	(0.000)	0.883	(0.000)	0.884	(0.000)	0.882	(0.000)	0.878	(0.000)	0.837	(0.000)
$ExA_1(3)$	0.759	(0.000)	0.753	(0.000)	0.757	(0.000)	0.756	(0.000)	0.761	(0.000)	0.754	(0.000)
$ExA_3(3)$	0.980	(0.000)	0.971	(0.000)	0.966	(0.000)	0.923	(0.000)	0.888	(0.000)	0.841	(0.000)

Panel B: Correlation with the Slope Factor

Maturity	3-month		6-month		1-year		2-year		5-year		10-year	
EGARCH	-0.407	(0.000)	-0.371	(0.000)	-0.381	(0.000)	-0.284	(0.000)	-0.219	(0.000)	-0.001	(0.980)
$CA_1(3)$	0.038	(0.476)	0.041	(0.437)	0.040	0.4526	0.041	(0.441)	0.038	(0.472)	0.041	(0.439)
$CA_2(3)$	-0.266	(0.000)	-0.272	(0.000)	-0.274	(0.000)	-0.272	(0.000)	-0.215	(0.000)	-0.092	(0.080)
$CA_3(3)$	-0.651	(0.000)	-0.625	(0.000)	-0.556	(0.000)	-0.443	(0.000)	-0.265	(0.000)	-0.087	(0.010)
$EA_1(3)$	0.027	(0.610)	0.032	(0.545)	0.029	0.584	0.030	(0.572)	0.026	(0.629)	0.032	(0.551)
$EA_2(3)$	-0.264	(0.000)	-0.269	(0.000)	-0.271	(0.000)	-0.269	(0.000)	-0.217	(0.000)	-0.100	(0.060)
$ExA_1(3)$	0.025	(0.637)	0.031	(0.563)	0.028	0.602	0.028	(0.595)	0.024	(0.651)	0.030	(0.572)
$ExA_3(3)$	-0.672	(0.000)	-0.638	(0.000)	-0.552	(0.000)	-0.421	(0.000)	-0.260	(0.000)	-0.108	(0.040)

Panel C: Correlation with the Curvature Factor

Maturity	3-month		6-month		1-year		2-year		5-year		10-year	
EGARCH	-0.462	(0.000)	-0.418	(0.000)	-0.423	(0.000)	-0.321	(0.000)	-0.2869	(0.000)	-0.129	(0.000)
$CA_1(3)$	-0.085	(0.100)	-0.081	(0.123)	-0.083	(0.117)	-0.082	(0.121)	-0.0846	(0.109)	-0.081	(0.123)
$CA_2(3)$	-0.492	(0.000)	-0.501	(0.000)	-0.504	(0.000)	-0.502	(0.000)	-0.427	(0.000)	-0.266	(0.000)
$CA_3(3)$	-0.653	(0.000)	-0.662	(0.000)	-0.651	(0.000)	-0.617	(0.000)	-0.466	(0.000)	-0.253	(0.000)
$EA_1(3)$	-0.091	(0.084)	-0.086	(0.105)	-0.089	(0.093)	-0.088	(0.097)	-0.0921	(0.081)	-0.086	(0.103)
$EA_2(3)$	-0.489	(0.000)	-0.497	(0.000)	-0.500	(0.000)	-0.498	(0.000)	-0.4298	(0.000)	-0.276	(0.103)
$ExA_1(3)$	-0.095	(0.073)	-0.089	(0.092)	-0.092	(0.080)	-0.091	(0.083)	-0.0956	(0.070)	-0.089	(0.090)
$ExA_3(3)$	-0.679	(0.000)	-0.686	(0.000)	-0.662	(0.000)	-0.613	(0.000)	-0.479	(0.000)	-0.285	(0.000)

Notes: This table reports correlations between the model-implied conditional volatility and the common factors affecting the co-movements of yields as well as correlations between the EGARCH volatility estimates on changes in yields and the common factors affecting the co-movements of yields . The level factor is defined as the 3-month yield. The slope is measured as the difference between the 10-year yield and the 3-month yield. The curvature corresponds to the yield on a butterfly that is long in the 10-year and the 3-month maturity bonds and short in two 1-year maturity bonds. Asymptotic p-values, computed using 5 Newey and West lags, are given in parentheses.

Table 9: Parameters Estimates for the Essentially Affine $A_1(3)$ Model.
Separate Estimation for Short-Maturity and Long-Maturity Bonds

	Panel A: Short Maturities			Panel B: Long Maturities		
	Factor			Factor		
	1	2	3	1	2	3
δ_0	0.0246 (0.0026)			0.0388 (0.0075)		
δ_{1j}	0.0072 (0.0011)	0.0007 (0.0001)	0.0037 (0.0005)	0.0058 (0.0046)	0.0057 (0.0047)	-0.0021 (0.0030)
κ_{1j}	0.5441 (0.0188)	0	0	0.0171 (0.0250)		
κ_{2j}	-0.0499 (0.0613)	1.8986 (0.1215)	40.0197 (3.0456)	0.1726 (0.1042)	0.6113 (0.0651)	0.3588 (0.2221)
κ_{3j}	0.7438 (0.0297)	0.0055 (0.1213)	9.3885 (0.4718)	0.7553 (0.3365)	0.1299 (0.1050)	1.0864 (0.1627)
θ_j	4.2491 (1.1630)	0	0	0.4144 (0.3640)	0	0
λ_{0j}	-0.622 (0.1442)	0.1353 (0.0395)	-0.3529 (0.7022)	-0.0555 (0.0364)	-1.5567 (2.1220)	0.1642 (0.2426)
$\lambda_{1(1j)}$	0	0	0	0	0	0
$\lambda_{1(2j)}$	-7.3524 (2.6496)	-0.1399 (0.1285)	22.7333 (10.1124)	4.7561 (0.4965)	-0.1931 (0.1340)	1.0456 (0.5644)
$\lambda_{1(3j)}$	-0.1279 (0.1301)	-0.3054 (0.5661)	-10.8474 (2.8236)	-0.6262 (0.5033)	-0.1738 (0.3418)	-0.9059 (0.3991)
β_1	1	0	0	1	0	0
β_2	39.5981 (12.5190)	0	0	2.5237 (0.9512)	0	0
β_3	0.5962 (0.0778)	0	0	0.1856 (0.2638)	0	0
L_{1j}	0.0027 (0.0001)	0	0	0.0022 (0.0001)	0	0
L_{2j}	-0.0001 (0.0001)	0.0008 (0.0000)	0	-0.0004 (0.0001)	0.0009 (0.0000)	0
L_{3j}	-0.0003 (0.0002)	-0.0001 (0.0000)	0.0008 (0.0000)	-0.0008 (0.0001)	-0.0001 (0.0000)	0.0013 (0.0001)
Log Likelihood	10802.00			10660.84		

Notes: In Panel A, we estimate the essentially affine $A_1(3)$ model on yields with a maturity of a less than one year. Yields with maturities of 1 month, 5 months and 9 months are assumed to be measured exactly, whereas yields with maturities of 3 months, 7 months and 1 year are assumed to be measured with errors. In Panel B, we estimate the essentially affine $A_1(3)$ model on yields with a maturity of a more than one year. Yields with maturities of 1 year, 3 years and 7 years are assumed to be measured exactly, whereas yields with maturities of 2 years, 5 years and 10 years are measured with errors. Zero-coupon yields are interpolated using the unsmoothed Fama and Bliss method. Standard errors are given in parentheses.

Table 10: Regressions of EGARCH Volatility Estimates of Changes in Yields on the Volatility Implied by the Essentially Affine $A_1(3)$ Model

Panel A: Short Maturities

Maturity	1-month	3-month	5-month	7-month	9-month	1-year
Intercept	-0.0058	-0.0034	-0.0031	-0.0031	-0.0030	-0.0021
	(-4.0905)	(-2.9962)	(-3.0656)	(-2.9504)	(-2.8795)	(-2.2472)
Slope	1.9087	1.6088	1.4731	1.4312	1.3923	1.2631
	(7.0799)	(6.6892)	(7.4423)	(6.9551)	(7.0169)	(7.0223)
Correlation	0.7112	0.6731	0.7159	0.7155	0.7148	0.7233

Panel B: Long Maturities

Maturity	1-year	2-year	3-year	5-year	7-year	10-year
Intercept	-0.0056	-0.0027	-0.0022	-0.0017	-0.0031	-0.0004
	(-3.3775)	(-2.2570)	(-2.2556)	(-2.9934)	(-4.5076)	(-1.0082)
Slope	1.7858	1.5281	1.4599	1.4225	1.7655	1.0833
	(5.9833)	(5.6799)	(6.3316)	(9.5968)	(9.7334)	(10.4407)
Correlation	0.6697	0.6605	0.6991	0.7760	0.7512	0.7426

Notes: We regress conditional volatility implied by EGARCH estimates of changes in yields on the conditional volatility implied by the essentially affine $A_1(3)$ Model. The model-implied volatility is estimated using yields with a maturity less than one year in Panel A, and using yields with a maturity more than one year in Panel B. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.

Table 11: In- and Out-of-Sample RMSE

Panel A: In Sample RMSE of the Change and the Level of Yields

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
Change in Yields	32.1970	27.2600	21.8480	15.6870	7.0358	5.5182
Level of Yields	32.0170	26.5740	21.8070	14.6740	6.9509	4.0134

Panel B: Out-of-Sample RMSE of the Change and the Level of Yields

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
Change in Yields	22.6460	13.6970	9.9599	8.0157	7.8069	9.5082
Level of Yields	20.8020	12.6820	9.7908	7.5843	6.9057	7.5711

Notes: We use the QML estimates from Table 2 to compute in- and out-of-sample root mean squared errors (RMSE) (in basis points) of the model implied conditional volatility with respect to EGARCH volatility estimates. The in-sample period is January 1970-December 1999, and the out-of-sample period is January 2000-December 2003.

Table 12: Regressions of EGARCH Volatility Estimates of Changes in Yields on the Instantaneous Volatility of the Short Rate Implied by the Essentially Affine $A_1(3)$ Model

Panel A: Estimation using Short-Maturity and Long-Maturity Yields

Maturity	1 Month	3 Months
Intercept	-0.0034	-0.0046
	(-1.8836)	(-2.7319)
Slope	1.3817	1.4837
	(4.4181)	(5.1692)
Correlation	0.5671	0.5987

Panel B: Estimation using Short-Maturity Yields Only

Maturity	1 Month	3 Months
Intercept	-0.0031	-0.0035
	(-2.7480)	(-3.0657)
Slope	1.4781	1.4541
	(6.6039)	(6.7079)
Correlation	0.6871	0.6645

Notes: We regress of conditional volatility implied by EGARCH estimates of changes in yields based on 1-month and 3-month yields on the instantaneous volatility of the short rate implied by the essentially affine $A_1(3)$ Model. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.

Table 13: Regressions of Realized Volatility Based Estimates on Model Implied and EGARCH volatility

Panel A: ARMA(1,1) Realized Variance Model

	Maturity					
Regressor	**3-month**	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
$EA_1(3)$	0.6204	0.6345	0.5978	0.6290	0.6041	0.5658
EGARCH(1,1)	0.9194	0.9245	0.8914	0.8909	0.8148	0.7247

Panel B: Ex-Post Realized Variance Model

	Maturity					
Regressor	**3-month**	**6-month**	**1-year**	**2-year**	**5-year**	**10-year**
$EA_1(3)$	0.5504	0.5859	0.5179	0.5408	0.5268	0.4299
EGARCH(1,1)	0.8586	0.8827	0.8212	0.8170	0.7207	0.5522

Notes: Panel A reports on the correlations between the conditional variance forecast implied by the realized variance model on the one hand and either EGARCH volatility or the essentially affine $A_1(3)$ model on the other hand. Panel B reports on the correlation between the ex-post realized volatility and either EGARCH volatility or the essentially affine $A_1(3)$ model

Table 14: Parameters Estimates and Regression Results using Different Assumptions on Measurement Errors

Panel A: Results from Regressions of EGARCH Volatility Estimates on the Conditional Volatility Implied by the Essentially Affine $A_1(3)$ Model Estimated with QML.

Maturity	6-month	1-year	3-year	5-year	7-year	10-year
Intercept	-0.0058	-0.0035	-0.0023	-0.0017	-0.0031	-0.0004
	(-2.9623)	(-2.5360)	(-2.2810)	(-2.9933)	(-4.6165)	(-1.0123)
Slope	1.6768	1.5698	1.4665	1.4380	1.7783	1.0812
	(5.1352)	(5.6918)	(6.1759)	(9.4334)	(9.9829)	(10.9204)
Correlation	0.6091	0.6502	0.6916	0.7754	0.7613	0.7461

Panel B: Results from Regressions of EGARCH Volatility Estimates on the Conditional Volatility Implied by the Essentially Affine $A_1(3)$ Model Estimated using the Extended Kalman Filter.

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
Intercept	-0.0044	-0.0030	-0.0026	-0.0008	0.0000	0.0000
	-2.6881	-2.2981	-2.3047	-0.9457	-0.1287	-0.0552
Slope	1.4869	1.4449	1.4297	1.1388	0.9757	0.9898
	5.1805	5.4595	6.0397	5.8153	10.1492	13.6566
Correlation	0.5994	0.6398	0.6794	0.6719	0.8317	0.8268

In Panel A, we regress model implied conditional volatility on EGARCH estimates of changes in yields. Model-implied volatility is obtained by estimating the model using QML on a sample of monthly data from January 1970 to December 1999. Yields with maturities of 1 year, 5 years and 10 years are assumed to be measured exactly, whereas yields with maturities of 6 months, 3 years and 7 years are measured with errors. Zero-coupon yields are interpolated using the unsmoothed Fama and Bliss method. Panel B reports the results of a similar regression, where model-implied volatility is obtained by estimating the model using the extended Kalman filter on a sample of monthly yields from January 1970 to December 1999, with maturities of 3 months, 6 months, 1 year, 2 years, 5 years and 10 years. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.

Table 15: The McCulloch and Kwon Cubic Spline Method

Panel A: Regressions of EGARCH Volatility Estimates of Changes in Yields on the Conditional Volatility Implied by the Essentially Affine $A_1(3)$ Model Estimated on the Sample Period 1970-1999.

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
Intercept	-0.0056	-0.0045	-0.0040	-0.0040	-0.0008	0.0000
	(-3.0071)	(-2.9816)	(-2.8223)	(-2.8223)	(-1.4895)	(0.1128)
Slope	1.6760	1.7465	1.6901	1.6901	1.1856	0.9663
	(5.1516)	(5.6061)	(5.7421)	(5.7421)	(8.5468)	(10.1192)
Correlation	0.6013	0.6527	0.6631	0.6631	0.8040	0.7371

Panel B: Regressions of EGARCH Volatility Estimates of Changes in Yields on the Conditional Volatility Implied by the Essentially Affine $A_1(3)$ Model Estimated on the Sample Period 1952-1999.

Maturity	3-month	6-month	1-year	2-year	5-year	10-year
Intercept	-0.0033	-0.0018	-0.0019	-0.0012	-0.0019	-0.0012
	(-2.9677)	(-2.2039)	(-2.5933)	(-2.4103)	(-4.4089)	(11.7242)
Slope	1.3223	1.2605	1.3402	1.2654	1.5655	1.4424
	(5.8864)	(6.2287)	(7.0964)	(8.7323)	(10.9046)	(11.7242)
Correlation	0.5954	0.6185	0.6719	0.7468	0.8166	0.8030

In Panel A, we report regression results for volatility implied by the essentially affine $A_1(3)$ model estimated using QML on a sample of monthly data from January 1970 to December 1999. Yields with maturities of 6 months, 2 years and 10 years are assumed to be measured exactly, whereas yields with maturities of 3 months, 1 year and 5 years are measured with errors. Zero-coupon yields are interpolated using the McCulloch and Kwon cubic spline method. Panel B reports regressions results using assumptions identical to Panel A, except that the sample period is January 1952 to December 1999. Asymptotic t-statistics, computed using five Newey and West lags, are reported in parentheses.