Can Affine Models Capture the Dynamics of Risk Premia and Volatility in Bond Yields? *

Peter Feldhütter †

This draft: December 30, 2006
First draft: March 1, 2006

Abstract

This paper estimates 3-factor extended and essentially affine models and asks whether 3-factor extended models better capture time-varying risk premia and volatility. The answer with regards to risk premia is yes: risk premia in extended affine models do match historical risk premia better and the improvement increases with the number of stochastic volatility factors. Regarding time-varying volatility extended models capture volatility slightly better but none of the models capture the high volatility during the Fed experiment. Also, it is shown that extended models match the distribution of yields better than essentially affine models, but the Feller condition limits the ability of the extended model with one stochastic factor to generate strongly skewed and fat-tailed distributions of long-maturity yields under the risk-neutral measure.

*I would like to thank Gregory Duffee, Scott Joslin, Jesper Lund, Mads Stenbo Nielsen, Kenneth Singleton, Anders Bjerre Trolle, and seminar participants at a Berkeley graduate student seminar for helpful comments.

†Department of Finance, Copenhagen Business School, Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark (e-mail: pf.fi@cbs.dk).

1 Introduction

Empirical evidence suggests that risk premia and volatility in U.S. interest rates are time-varying. The excess return on a bond - the return of the bond in one period minus the short interest rate - is documented to be time-varying and positively related to the slope of the yield curve. It is also documented that volatility of yields is time-varying and positively related to the level of yields. Using more than 50 years of U.S. Treasury bond data, this paper asks the question whether these facts of U.S. interest rates along with the first four moments of yields can be better captured by 3-factor extended than 3-factor essentially affine term structure models. The two types of affine models are the most popular term structure models due to their analytic tractability and rich risk premium specifications, but while essentially affine models have been compared to completely affine and quadratic affine models in earlier literature, little is known about the properties of extended affine models relative to essentially affine models.

A general characterization of N-factor affine models is given by Dai and Singleton (2000) who decompose the N factors into m stochastic volatility factors and $N-m$ factors not entering volatility (denoted $A_m(N)$) and propose completely affine models by letting risk premia be proportional to the volatility of the state variables. The essentially affine model of Duffee (2002) lets the risk premia of non-volatility factors be affine functions of all factors but evidence in Dai and Singleton (2002) and Duffee (2002) suggests that essentially affine models might fit the first or second moment of yields but not both. The extended affine model of Cheridito, Filipovic, and Kimmel (2006) allows the risk premium of each volatility factor to be an affine function of all volatility factors and their risk premium specification has the potential to solve the volatility-risk premium tension in essentially affine models since the tight restrictions on the risk premium of volatility factors are relaxed[1]. However, there are two reasons why extended affine models might fail this task. First, the risk premium of a volatility factor cannot depend on the non-volatility factors and the dependence on other volatility factors is restricted because correlations between volatility processes have to be positive. Second, the (multivariate) Feller restriction is imposed under P and Q which restricts the dynamics of volatility factors. Because of the Feller condition

[1] Collin-Dufresne, Goldstein, and Jones (2005) also propose this modification of risk premia.

the extended affine models do not nest neither essentially nor completely affine models.

Campbell and Shiller (1991) show that the regression coefficient of yield changes regressed on the slope of the term structure is negative and increasingly so with maturity implying that excess returns are positively related to the slope of the term structure. This result contradicts the Expectation Hypothesis which predicts a coefficient of one if risk premia are constant. In addition, a model's risk premium specification can be tested by adjusting the Campbell-Shiller regressions for model-implied risk premia such that coefficients of 1 are restored as shown in Dai and Singleton (2002). Using the two sets of regression coefficients as measuring sticks the results in this paper show that extended models capture risk premia better than essentially affine models. Although none of the extended and essentially affine models with stochastic volatility match the $A_0(3)$ model, the three extended models with stochastic volatility outperform their essentially affine counterparts and the difference increases with the number of volatility factors. In essentially affine models the ability to capture time-varying risk premia decreases strongly with the number of volatility factors but in the extended models there is not a large difference between models with 1, 2, and 3 volatility factors and all three extended models do as well or better than the essentially affine $A_1(3)$ model.

The level effect in volatility is documented by among others Chapman and Pearson (2001) and this effect can be measured by regressing squared yield changes on the level, slope, and curvature of the yield curve and finding a positive and highly significant coefficient on the level of the yield curve. All models predict a positive level effect but none of the models capture the magnitude of the effect or simultaneously match the correct sign on the slope and curvature coefficients. Brandt and Chapman (2003) conclude that quadratic term structure models perform better than essentially affine models and point to these regression results as the most important factor in the difference of fit. However, comparing model-implied volatility with historic volatility I show that both model classes capture time-varying volatility quite well with exception of the volatile Fed experiment period in the beginning of the 80's, and the $A_1(3)$ extended and essentially affine models largely match the size and sign of the level, slope, and curvature coefficients for the subperiods before and after the Fed experiment. The results also show that extended models capture time-varying volatility slightly better than essentially affine models although the ability decreases with the number of stochastic volatility factors for both models.

Both extended and essentially affine models capture the first two unconditional moments of yields at all maturities, but the moments are estimated with more precision in extended models. More importantly, the distribution of yields in extended models are more in accordance with the historical distribution than in essentially affine models: skewness and kurtosis of yields in extended models are much closer to historical skewness and kurtosis. The intuition behind this result is clear when examining the $A_1(3)$ model. Empirically, skewness and kurtosis of yields with long time to maturity depend largely on a single parameter and while this parameter is shared by the actual and risk-neutral dynamics in the essentially affine $A_1(3)$ model it is allowed to take different values under the two measures in the extended counterpart. In the essentially affine model, the historical distribution of long yields 'inherits' a skewed and fat-tailed model skewness and kurtosis can differ such that the historical distribution is better captured.

While the distribution of yields in extended models is more in accordance with the historical distribution, I show that the essentially affine $A_1(3)$ model generates more skewed and fat-tailed risk-neutral distributions than the extended $A_1(3)$ model and argue that due to the Feller restriction the extended $A_1(3)$ model cannot generate the skewness and kurtosis that the data calls for without sacrificing the cross-sectional fit of yields. Therefore, the extended $A_1(3)$ might not price instruments that are sensitive to the tail-behavior of yields well - such as out-of-the-money options.

Related to this paper is Cheridito, Filipovic, and Kimmel (2006) who estimate and compare 1, 2, and 3-factor extended and essentially affine models. However, they impose the Feller condition in all models and restrict some of the volatility parameters in the $A_1(3)$ and $A_2(3)$ models to be zero. In addition, they only look at the first two unconditional moments of yield changes. In another related paper, Almeida, Graveline, and Joslin (2006) examine whether the inclusion of options in estimation of extended affine models helps predict excess returns and volatility. While they focus on the effect of observing options in extended models, I focus on comparing the non-nested essentially and extended affine models.

The paper is organized as follows. Section 2 describes the features in the U.S. term structure that affine models should match. The affine framework is set up in section 3 and the estimation methodology MCMC is explained in 4. Results are presented in section 5 and section 6 concludes.

2 Features of the U.S. Term Structure

In this paper 3-factor essentially and extended affine term structure models are compared across a set of U.S. term structure 'features' that are easily interpretable and has proven difficult for term structure models to match. Earlier studies have used a subset of the features to compare completely affine, essentially affine, and/or quadratic term structure models. This study tries to answer whether extended affine models can overcome the inability of completely/essentially affine models to match some of these features.

I use month-end (continuously compounded) 1, 2, 3, 4, and 5 years zero-coupon yields extracted from US Treasury security prices by the method of Fama and Bliss (1987). The data are from the Center for Research in Security Prices and cover the period 1952:6 to 2004:12. The data is discussed in Elton and Green (1998) who suggest that CRSP bond prices contain more noise than industry-used prices, but higher quality data is not available for long sample periods. The data set is used both in this section to illustrate the U.S. term structure features and in the later estimation of affine models.

The yield curve is on average upward sloping.

The first feature of the U.S. term structure is that the unconditional mean of yields is rising with maturity. This is illustrated by taking the mean of yields across maturity which is done in Table 1. On average the five-year yield is 59 basis points higher than the one-year yield and it is higher in 79 % of the months.

[Table 1 about here.]

Expected excess returns are time-varying and positively related to the slope of the yield curve.

Expected excess returns in U.S. Treasury bonds vary across time and maturity. This well-established fact is documented in a series of papers starting with Fama (1984) and is stable across time periods and data sets. Expected excess returns tend to be positive when the slope of the yield curve is steep and negative when the yield curve is flat or downward-sloping and the effect increases with the maturity of the bond. Campbell and Shiller (1991) document this phenomenon by regressing future yield changes (inversely related to excess returns) on the current slope of the yield curve. Specifically, the regression is

$$Y(t+1, n-1) - Y(t, n) = const + \phi_n[\frac{Y(t, n) - Y(t, 1)}{n-1}] + res, \tag{1}$$

where $Y(t, n)$ is the n-year zero-coupon yield at time t and for the data set in this paper the coefficients are given in Table 2. Dai and Singleton (2002) call the regression $LPY(i)$. With constant risk premia the expectation theory predicts that the coefficients equal 1, but the actual coefficients are negative and increasingly so with maturity.

[Table 2 about here.]

Dai and Singleton (2002) show that the regression coefficients of one in the Campbell-Shiller regressions are restored by adjusting for time-varying risk premiums and for completeness their derivation is restated in Appendix A. Dai and Singleton (2002) call this regression for $LPY(ii)$.

Volatility of yield changes is on average downward sloping for large maturities.

In contrast to the unconditional mean of yields, the unconditional volatility - defined as standard deviation - of yield changes is falling with maturity. As illustrated in Table 1 the monthly volatility falls from 49.3 basis points for the one-year yield to 36.2 basis points for the 5-year yield. This phenomenon is not consistent for short maturities over different time periods since the volatility curve is hump-shaped when using data only from the Greenspan era 1987:8-2004:12 but it seems to be consistent for maturity 2-3 years and more[2]. However, for the sample period used in this paper, the volatility curve is downward-sloping for all maturities.

Volatilities of yields are time-varying and positively related to the level of yields.

Yield volatility has historically been time-varying and positively correlated with interest rates, which is documented in Brandt and Chapman (2003) and Piazzesi (2003). This is seen in the data set by regressing squared monthly yield changes on the level, slope, and curvature of the yield curve -

[2]See Piazzesi (2005) for a detailed discussion.

the three components identified by Litterman and Scheinkman (1991) that explain most return variability across the maturity spectrum[3]. Table 3 shows the regression coefficients for the data and the level factor is strongly significant across all maturities. The table also shows that the relation between volatility and curvature is positive (although mostly insignificant). This is consistent with Christiansen and Lund (2005) who argue that curvature measures the cost of convexity and this cost is high when volatility is high. Finally, the table shows an insignificant but mostly negative relation between slope and volatility. This is slightly surprising since the slope of the yield curve depends on the risk premium for the long-maturity bond and therefore a positive relationship is expected but as discussed in section 5.3 the negative regression coefficients are due to the combination of high volatility and on average inverted yield curves during the Fed experiment 1979-1982 (see also Christiansen and Lund (2005)).

[Table 3 about here.]

To sum up, essentially affine and extended affine models are compared by matching averages and regression coefficients found in the actual U.S. term structure data with model-implied averages and coefficients.

3 Affine Term Structure Models

In this section affine term structure models are characterized using the Dai and Singleton (2000) framework.

3.1 Bond Pricing

The short rate r_t is an affine vector of unobserved state variables $X_t = (X_t^1, ..., X_t^N)'$,

$$r_t = \delta_0 + \delta_x' X_t, \tag{2}$$

and X_t follows an affine diffusion,

$$dX_t = (K_0^Q - K_1^Q X_t)dt + \Sigma\sqrt{S_t}d\widetilde{W}_t, \tag{3}$$

[3]Regressing on level, slope, and curvature instead of yields directly strongly reduces the correlation of the regressors.

where $\widetilde{W}_t$ is an N-dimensional standard Brownian motion under Q, K_0^Q is a vector of length N while K_1^Q, Σ, and S_t are $N \times N$ matrices. S_t is a diagonal matrix with elements $[S_t]_{ii} = \alpha_i + \beta_i' X_t$, where α_i is a scalar while β_i is an N-vector. Parameter restrictions ensuring that the dynamics of X_t are well-defined are given in Dai and Singleton (2000).

Duffie and Kan (1996) show that bond prices are exponential-affine

$$P(t,\tau) = e^{A^*(\tau) - B^*(\tau)' X_t},$$

where $P(t,\tau)$ denotes the price of a zero coupon bond at time t that matures at time $t+\tau$ and the functions $A^*(\tau)$ and $B^*(\tau)$ solve the ODEs

$$\frac{dA^*(\tau)}{d\tau} = -K_0^{Q'} B^*(\tau) + \frac{1}{2}\sum_{i=1}^{N} [\Sigma' B^*(\tau)]_i^2 \alpha_i - \delta_0, \tag{4}$$

$$\frac{dB^*(\tau)}{d\tau} = -K_1^{Q'} B^*(\tau) - \frac{1}{2}\sum_{i=1}^{N} [\Sigma' B^*(\tau)]_i^2 \beta_i + \delta_x. \tag{5}$$

The corresponding (continuously compounded) yield of bond $P(t,\tau)$ is

$$Y(t,\tau) = A(\tau) + B(\tau) X_t,$$

where $A(\tau) = \frac{-A^*(\tau)}{\tau}$ and $B(\tau) = \frac{B^*(\tau)'}{\tau}$.

I adopt the normalizations in the canonical form of Dai and Singleton (2000) and the restrictions for all 3-factor models are given in Appendix B.

3.2 Risk Premia

The stochastic discount factor M determining Q can be written as

$$\frac{dM_t}{M_t} = -r_t dt - \Lambda_t' dW_t,$$

where W_t is a Brownian motion under the actual measure P. The dynamics of X_t under P is given as

$$dX_t = (K_0^Q - K_1^Q X_t)dt + \Sigma S_t^{\frac{1}{2}} \Lambda_t dt + \Sigma S_t^{\frac{1}{2}} dW_t$$

and Dai and Singleton (2000) choose the *completely affine* market price of risk as

$$S_t^{\frac{1}{2}} \Lambda_t = S_t \Phi_1$$

and all variation in the price of risk vector is then due to variation in S_t.

Duffee (2002) proposes an *essentially affine* market price of risk

$$S_t^{\frac{1}{2}}\Lambda_t = S_t\Phi_1 + I^-\Phi_2 X_t, \tag{6}$$

where I^- is an $N \times N$ diagonal matrix with $I^-_{ii} = 1_{\{\inf(\alpha_i+\beta_i' X_t)>0\}}$ and Φ_2 is a $N \times N$ matrix. The essentially affine market price of risk nests the completely affine and extends the flexibility of the price of risk of the $N - m$ non-volatility factors.

Cheridito, Filipovic, and Kimmel (2006) propose an *extended affine* price of risk

$$S_t^{\frac{1}{2}}\Lambda_t = \lambda_1 + \lambda_2 X_t, \tag{7}$$

where λ_1 is an N-vector and λ_2 is an $N \times N$ matrix that possibly has restrictions ensuring that the process X is well defined under P. Compared to essentially affine models their specification adds flexibility to the price of risk of the m volatility factors without restricting the flexibility in the price of risk of the $N - m$ non-volatility factors. However, the flexibility comes at a cost. To avoid arbitrage opportunities the volatility matrix S_t must be strictly positive and therefore the parameter vector has to satisfy the multivariate generalization of the Feller condition. As a consequence the extended affine models do not nest neither the essentially nor the completely affine models.

In Appendix B the parametrization of all 3-factor essentially and extended affine models is explained in detail.

4 Estimation

In estimating the affine models I adopt a Bayesian approach and estimate the models by MCMC as proposed by Eraker (2001)[4]. The approach has several advantages which are useful for the following analysis. First of all, every yield can be observed with error. Usually, it is assumed that three

[4]For a general introduction to MCMC see Robert and Casella (2004) and for a survey of MCMC methods in financial econometrics see Johannes and Polson (2003). Examples of estimating affine term structure models in a single-factor setting are Mikkelsen (2001) and Sanford and Martin (2005) while multi-factor examples are Lamoureux and Witte (2002) and Bester (2004).

yields are observed without error such that state variables can be extracted from yields[5]. Second, the models can be estimated without imposing additional parameter restrictions not implied by the theoretical model. Dai and Singleton (2002) restrict the volatility factors in the $A_2(3)$ and $A_3(3)$ to be independent. Cheridito, Filipovic, and Kimmel (2006) restrict the volatility of the non-volatility factors to be constant, thereby restricting the two β parameters in the $A_1(3)$ and $A_2(3)$ models to be zero. Since I am interested in comparing models both across different risk premium specifications and across the number of volatility factors, I do not want to disadvantage some models compared to others by imposing parameter restrictions. Third, the main interest in the analysis is whether the models can capture the size and sign of certain regression coefficients obtained by running the regressions on the actual data. MCMC facilitates the construction of the marginal density of any function of the parameters and state variables and therefore the marginal density of any regression coefficient of interest can be obtained taking into account uncertainty about parameters and state variables[6]. Fourth, MCMC can easily handle parameter restrictions, while optimization algorithms of traditional frequentist methods often perform poorly in presence of hard parameter constraints[7].

4.1 Estimating Affine Term Structure Models

At time $t = 1, ..., T$ k yields are observed and they are stacked in the k-vector $Y_t = (Y(t, \tau_1), ..., Y(t, \tau_k))'$. The yields are all observed with a measurement error

$$Y_t = A + BX_t + \epsilon_t$$

where A is a k-vector and B a $k \times N$ matrix. I assume that the measurement errors are independent and normally distributed with zero mean and common variance such that

$$\epsilon_t \sim N(0, D), \qquad D = \sigma^2 I_k.$$

[5]See for example Dai and Singleton (2002), Duffee (2002), Cheridito, Filipovic, and Kimmel (2006).

[6]An example of this approach is found in Lamoureux and Witte (2002) who find the density of coefficients of an "Expectation Hypothesis" regression from 2- and 3-factor CIR models with independent factors.

[7]See for example Cheridito, Filipovic, and Kimmel (2006).

I could allow for a more flexible parametrization of the error variances, letting for example each measurement error have its own variance. However, the effect of individual variances is that in the MCMC estimation of some of the models the shortest yield is fit almost perfectly - an effect which is difficult to interpret from an economic point of view[8].

The parameters of the model and the variances of the measurement errors are stacked in the vector $\Phi = (K_0^Q, K_1^Q, \beta, \lambda, \delta, D)$. In the estimation the latent variables (X_t) are treated as parameters but for clarity they are separated in the vector X.

I am interested in samples from the target distribution $p(\Phi, X|Y)$. The Hammersley-Clifford Theorem (Hammersley and Clifford (1970) and Besag (1974)) implies that samples are obtained from the target distribution by sampling from the full conditionals

$$\begin{gathered} p(K_0^Q|K_1^Q, \beta, \lambda, \delta, D, X, Y) \\ p(K_1^Q|K_0^Q, \beta, \lambda, \delta, D, X, Y) \\ \vdots \\ p(X|K_0^Q, K_1^Q, \beta, \lambda, \delta, D, Y) \end{gathered}$$

so MCMC solves the problem of simulating from the complicated target distribution by simulating from simpler conditional distributions. Specifically, draw $i+1$ of the parameters $(K_0^Q, K_1^Q, \beta, \lambda, D, X)$ in the MCMC algorithm is obtained by drawing from the full conditionals

$$\begin{gathered} p(K_0^Q|(K_1^Q)_i, \beta_i, \lambda_i, \delta_i, D_i, X_i, Y) \\ p(K_1^Q|(K_0^Q)_{i+1}, \beta_i, \lambda_i, \delta_i, D_i, X_i, Y) \\ \vdots \\ p(X|(K_0^Q)_{i+1}, (K_1^Q)_{i+1}, \beta_{i+1}, \lambda_{i+1}, \delta_{i+1}, D_{i+1}, Y). \end{gathered}$$

If one samples directly from a full conditional the resulting algorithm is the Gibbs sampler (Geman and Geman (1984)). If it is not possible to sample directly from the full conditional distribution one can sample by using the Metropolis-Hastings algorithm (Metropolis et al. (1953)). I use a hybrid MCMC algorithm that combines the two since not all the conditional distributions are known.

---

[8] See Lamoureux and Witte (2002) for an example of this phenomenon in the estimation of a 3-factor model.

4.1.1 The Conditionals $p(X|\Phi)$ and $p(Y|\Phi, X)$

The conditional $p(X|\Phi)$ is used in several steps of the MCMC procedure and is calculated as

$$p(X|\Phi) = \Big(\prod_{t=1}^{T} p(X_t|X_{t-1}, \Phi)\Big) p(X_0).$$

The continuous-time specification in (3) is approximated using an Euler scheme (letting $\Sigma = I_N$)[9]

$$\begin{aligned} X_{t+1} &= X_t + \mu_t^P \Delta_t + \sqrt{\Delta_t S_t}\epsilon_{t+\Delta_t}, \\ \epsilon_{t+1} &\sim N(0, I_N), \end{aligned}$$

where Δt is the time between two observations and $\mu_t^P = K_0^Q - K_1^Q X_t + S_t^{\frac{1}{2}} \Lambda_t$ is the drift under P. Since S_t is diagonal

$$p(X|\Phi) \propto \prod_{i=1}^{N} \Big(\Big[\prod_{t=1}^{T} [S_{t-1}]_{ii}^{-\frac{1}{2}}\Big] \exp\big(-\frac{1}{2\Delta_t}\sum_{t=1}^{T}\frac{[\Delta X_t - \mu_{t-1}^P \Delta_t]_i^2}{[S_{t-1}]_{ii}}\big)\Big) p(X_0).$$

If the difference between the actual yields and the model implied yields at time t is denoted by $\hat{e}_t = Y_t - (A(\Phi) + B(\Phi)X_t)$, the density $p(Y|\Phi, X)$ can be written as

$$\begin{aligned} p(Y|\Phi, X) &\propto \prod_{i=1}^{k} \Big(D_{ii}^{-\frac{T}{2}} \exp(-\frac{1}{2D_{ii}} \sum_{t=1}^{T} \hat{e}_{t,i}^2)\Big) \\ &\propto \sigma^{-kT} \exp(-\frac{1}{2\sigma^2} \sum_{t=1}^{T} \hat{e}_t' \hat{e}_t) \end{aligned}$$

4.1.2 The Hybrid MCMC algorithm

According to Bayes' theorem the conditional of the risk premium parameters is given as

$$\begin{aligned} p(\lambda|\Phi_{\setminus\lambda}, X, Y) &\propto p(Y|\Phi, X) p(\lambda|\Phi_{\setminus\lambda}, X) \\ &\propto p(X|\Phi) p(\lambda|\Phi_{\setminus\lambda}) \end{aligned}$$

[9]The Euler scheme introduces some discretization error which may induce bias in the parameter estimates. This possible bias can be reduced using Tanner and Wong (1987)'s data augmentation scheme. However, Bester (2004), using also monthly yield data, report that data augmentation does not significantly affect parameter estimates. For the effect of data augmentation in a one-factor model see Sanford and Martin (2005).

where $\Phi_{\backslash\lambda}$ denotes the parameter vector without the parameters λ and it is used that $p(Y|\Phi, X)$ does not depend on λ. I assume that the priors are a priori independent and in order to let the data dominate the results a standard diffuse, noninformative prior, is adopted so $p(\lambda|\Phi_{\backslash\lambda}, X, Y) \propto p(X|\Phi)$ and the λ's can be Gibbs sampled one column at a time from a multivariate normal distribution.

The conditional of the variance of the measurement errors is given as

$$\begin{aligned} p(D|\Phi_{\backslash D}, X, Y) &\propto p(Y|\Phi, X)p(D|\Phi_{\backslash D}, X) \\ &\propto p(Y|\Phi, X)p(X|\Phi)p(D|\Phi_{\backslash D}) \\ &\propto p(Y|\Phi, X), \end{aligned}$$

since $p(X|\Phi)$ does not depend on D. σ^2 can therefore be Gibbs sampled from the inverse Wishart distribution, $\sigma \sim IW(\sum_{t=1}^{T} \hat{e}_t' \hat{e}_t, kT)$[10].

The conditional of the other model parameters is given as

$$\begin{aligned} p(\Phi_j|\Phi_{\backslash\Phi_j}, X, Y) &\propto p(Y|\Phi, X)p(\Phi_j|\Phi_{\backslash\Phi_j}, X) \\ &\propto p(Y|\Phi, X)p(X|\Phi)p(\Phi_j|\Phi_{\backslash\Phi_j}) \\ &\propto p(Y|\Phi, X)p(X|\Phi), \end{aligned}$$

which for none of the parameters K_0^Q, K_1^Q, β, and δ is a known distribution. To block sample the four sets of parameters I use the Random Walk Metropolis-Hastings algorithm (RW-MH). To sample Φ_j at MCMC step $i+1$, I propose Φ_j^{i+1} by drawing a multivariate normal distributed variable centered around Φ_j^i and accept it with probability $\min\left(1, \frac{f(\Phi_j^{i+1})}{f(\Phi_j^i)}\right)$ where f is the density $p(\Phi_j|\Phi_{\backslash\Phi_j}, X, Y)$.

The latent processes are sampled by sampling $X_t, t = 0, ..., T$ one at a time using the RW-MH procedure. For $t = 1, ..., T-1$ the conditional of X_t is given as

$$\begin{aligned} p(X_t|X_{\backslash t}, \Phi, Y) &\propto p(X_t|X_{t-1}, X_{t+1}, \Phi, Y_t) \\ &\propto p(Y_t|X_t, \Phi)p(X_t|X_{t-1}, X_{t+1}, \Phi) \\ &\propto p(Y_t|X_t, \Phi)p(X_t|X_{t-1}, \Phi)p(X_{t+1}|X_t, \Phi). \end{aligned}$$

For $t = 0$ the conditional is

$$\begin{aligned} p(X_0|X_1, \Phi, Y) &\propto p(X_1|X_0, \Phi, Y)p(X_0) \\ &\propto p(X_1|X_0, \Phi)p(X_0), \end{aligned}$$

[10]Equivalent to an inverted gamma distribution.

while for $t = T$ the conditional is

$$\begin{aligned} p(X_T|X_{\backslash X_T}, \Phi, Y) &\propto p(X_T|X_{T-1}, \Phi, Y) \\ &\propto p(Y_T|X_T, X_{T-1}, \Phi, Y_{\backslash Y_T})p(X_T|X_{T-1}, \Phi, Y_{\backslash Y_T}) \\ &\propto p(Y_T|X_T, \Phi)p(X_T|X_{T-1}, \Phi). \end{aligned}$$

Both the parameters and the latent processes are subject to constraints and if a draw is violating a constraint it can simply be discarded (Gelfand et al. (1992)). However, I use RW-MH to sample the risk premium parameters in the extended affine models since practically all the draws would otherwise be discarded due to the non-attainment parameter constraints. In estimating each model I use an algorithm calibration period of eight million draws, where the variances of the normal proposal distributions are set, a burn-in period of two million draws and an estimation period of four million draws. Due to lack of computer memory I keep every 200'th draw in the estimation period which leaves 20.000 draws[11]. Implementation details are given in Appendix C.

5 Results

In this section I compare affine models across the averages and regression coefficients defined in section 2 to see which models successfully capture the level and time-variability in mean and volatility. The parameter estimates are given in Table 4 through 7.

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[11]All random numbers in the estimation are draws from Matlab 7.0's generator which is based on Marsaglia and Zaman (1991)'s algorithm. The generator has a period of almost 2^{1430} and therefore the number of random draws in the estimation is not anywhere near the period of the random number generator.

There are several ways of calculating the coefficients for each model and the implications can vary strongly depending on how the coefficients are calculated. First, the coefficients can be calculated using fitted yields. Second, they can be derived analytically. Third, they can be calculated by simulating yields from the model. As Dai and Singleton (2002) note the dynamics of fitted yields is determined by both the model used in calculating them and the historic properties of yields, and therefore assessments of fit based on fitted yields can give very misleading impressions of the actual population distribution. Population moments and moments calculated from simulated yields (of length equal to the length of the data) may differ because of finite-sample biases. To see whether this is the case, I find the density of the Campbell-Shiller regression coefficients both in population and from simulated data. First the density of each population coefficient is obtained by analytically calculating the regression coefficient for each MCMC draw and empirically estimating a density based on the 20.000 analytical coefficients. Details of how to analytically calculate the regression coefficient are given in Appendix D. The second set of regression coefficients are calculated from simulated data. For every MCMC draw the regression coefficients are calculated by repeating a simulation of 631 months 100 times, calculating the regression coefficients for every draw, and taking the average regression coefficient over the 100 simulations. Ideally, this should be repeated for every MCMC draw to get the distribution of regression coefficients but since this is too time-consuming, this is done for every 50'th MCMC draw. This amounts to an average over 400 averages - averaging over a total of 40.000 simulations[12]. The results are given in Table 8.

[Table 8 about here.]

It is seen in the table that the finite-sample bias is small for some of the models (0.1 or less for the $A_0(3)$ model) but sizeable for others (0.5 or more for

[12] To assure that the difference is due to finite-sample bias and not the simulation method two checks is performed. First, I simulate once from every MCMC draw and average over the 20.000 simulations. This should give approximately the same coefficient estimates while giving larger confidence bands. The estimates for the $A_3(3)$ essentially affine model from this procedure only differs from the simulation method in the text on the third decimal. Second, I calculate the population coefficients using the estimated parameters and using the same parameters the average coefficients from 1.000 simulations are calculated. The difference between the model-implied coefficients and the coefficients from simulating is of the same sign and magnitude for the essentially affine $A_3(3)$ model as in Table 8.

the $A_2(3)$ and $A_3(3)$ essentially affine models). While the bias for all models go in the same direction, the magnitude is different and it is important to take this bias into account. For the rest of the paper the results are therefore based on simulated data using the aforementioned simulation procedure.

5.1 Risk Premia

Focusing now on the Campbell-Shiller regression coefficient instead of finite-sample bias, we see in Table 8 that the ability of the essentially affine models to capture the Campbell-Shiller regression coefficients decrease in the number of volatility factors. Only the $A_0(3)$ model is able to capture the downward sloping curve of CS coefficients and has confidence bands that contain the actual CS coefficients. All essentially affine models with stochastic volatility miss the slope and sign of the CS coefficients. This is consistent with findings in Dai and Singleton (2002). While the extended affine models are not able to beat the $A_0(3)$ model in terms of capturing the CS coefficients, they do better than their essentially affine counterparts. The coefficients of the extended affine $A_2(3)$ and $A_3(3)$ models are comparable to the coefficients of the best essentially affine model with stochastic volatility, $A_1(3)$, and the extended $A_1(3)$ matches the coefficients better than any of the essentially affine models with stochastic volatility.

The estimates of the risk-adjusted CS coefficients are given in Table 9. and the results are similar to those of the unadjusted CS coefficients. For essentially affine models the ability to capture the risk-adjusted coefficients of 1 decreases with the number of stochastic factors and only the $A_0(3)$ model captures the coefficients. All extended models do better than the essentially affine $A_1(3)$ model and the extended $A_1(3)$ has estimated risk-adjusted coefficients closer to 1 than the $A_2(3)$ and $A_3(3)$ extended models (except for the 5-year maturity).

[Table 9 about here.]

Figure 1 compares the essentially and extended affine models with the same number of stochastic volatility factors. The left panel shows the density of the 4-year CS coefficients while the right panel shows the density of the 4-year adjusted CS coefficients.

[Figure 1 about here.]

We see a sizeable improvement in time-varying predictability in the $A_2(3)$ model and a dramatic improvement in the $A_3(3)$ model. In the literature the essentially affine $A_0(3)$ and $A_1(3)$ models have generally been preferred over the $A_2(3)$ and $A_3(3)$ models partly because of the latter models' inability to capture time-varying risk-premia. The results with regards to risk premia indicate that the extended $A_1(3)$ model captures time-varying risk premia better than any other 3-factor essentially affine or extended model with stochastic volatility, but the difference between the extended $A_1(3)$, $A_2(3)$, and $A_3(3)$ models are much smaller than they are in the essentially affine models. Using 11 years of swap rates and cap data Almeida, Graveline, and Joslin (2006) show that the $A_1(3)$ and $A_2(3)$ extended models are able to capture the size and slope of the CS regression coefficients when options are included in estimation. In contrast, Duffee (2002) show that completely affine models cannot capture the time-varying risk premia and since the $A_3(3)$ completely and essentially affine models are identical, this suggests that at least the $A_(3)$ essentially affine model cannot capture the time-variability whether or not options are included. Altogether, this suggests that 3-factor extended models do have the flexibility to fit the CS regression coefficients although options might be needed in the estimation to fully capture the exact level, while the risk premium specification in essentially affine models, at least the $A_3(3)$ model and possibly the $A_2(3)$ model, is too restrictive to let the models capture the time-variability in risk premia even when other instruments are included in estimation.

5.2 Unconditional Mean and Volatility

Turning to the average of yields, Table 10 shows the estimated unconditional mean of yields in the affine models.

[Table 10 about here.]

All models capture the level of yields reasonably well but there are large differences in the precision of the different models. The extended models estimate the mean more precisely than the essentially affine. For example, the length of the confidence band for the 5-year yield in the essentially affine $A_1(3)$ model is 38.1 while it is 8.9 in the extended $A_1(3)$ model. While Table 10 shows estimates and confidence bands Figure 2 depicts the whole distribution in the models by showing box plots of the distribution of the 5-year yield[13].

[Figure 2 about here.]

The box plot shows that the distribution is right-skewed in all models with stochastic volatility and the skew is stronger for essentially affine models. This is due to the difference in risk premium specification in the models but a discussion of this is deferred until Section 5.4.

Extended models estimate the unconditional mean of yields more precisely and as it is seen in Table 11 they also estimate the unconditional volatility with larger precision than essentially affine models.

[Table 11 about here.]

All models capture the level of unconditional volatility but there is a large difference in the precision with which it is captured. The $A_0(3)$ model does not allow for time-varying volatility and therefore it can estimate the unconditional volatility with high precision. The confidence bands in the $A_0(3)$ model is a 4-6 times smaller than confidence bands in the extended affine models. While the extended models estimate the volatility with much less precision than the $A_0(3)$ model they are far better than the essentially affine models in terms of the size of confidence which are a factor 3-4 times smaller. For example, the confidence band for the 5-year volatility is 3.6 basis points wide in the $A_0(3)$ model, $17.4 - 27.5$ wide in the extended models, and $88.2 - 105.7$ wide in the essentially affine models. From the table we can therefore conclude that while all models statistically capture the level of unconditional volatility (meaning that the estimated coefficients in data are contained in the confidence bands) there is a very large difference in the precision with the $A_0(3)$ model having the tightest confidence bands and the essentially affine stochastic volatility models having the widest.

Figure 3 shows a box plot of the density of the unconditional 5-year volatility in the models. We see a similar pattern as in the unconditional mean: all distributions in models with stochastic volatility are right-skewed and the skew is stronger in essentially affine models than in extended models.

[Figure 3 about here.]

[13] In a box plot the box ends with the first and third quartile and has the median horizontally drawn inside. The "whiskers" extend to the farthest points that are within 1.5 times the interquartile range of the first and third quartiles, and dots are outliers.

5.3 Conditional Volatility

In section 2 a significant level effect in volatility is shown by regressing squared monthly yield changes on the level, slope, and curvature of the term structure, and if a model captures the dynamics of time-varying volatility correctly the model should replicate this coefficient. Table 12 shows the model-implied regression coefficients.

[Table 12 about here.]

The $A_0(3)$ model estimates the coefficient for every maturity to zero which is expected since the model does not accommodate stochastic volatility. All models with stochastic volatility capture the positive sign of the coefficients and the downward sloping curve of level coefficients with respect to maturity. The models largely agree on the size of the level coefficients although the extended $A_1(3)$ model estimates the coefficients somewhat lower than the other stochastic volatility models, but the actual size of the coefficients in the data is approximately twice as big as the model-implied coefficients, so none of the models capture the coefficients in the data. For example, the actual 3-year coefficient is 0.057 while in the affine models with stochastic volatility it is estimated to be in the range 0.022-0.033 and this difference is statistically significant according to the confidence bands of the model-implied coefficients.

In addition to the failure of replicating the correct level coefficients none of the models simultaneously replicate the correct sign of the slope and curvature regression coefficients. While the actual slope coefficients are negative (except for the 5-year maturity) and the curvature coefficients are positive, the models with 1 and 2 volatility factors predict positive coefficients on slope and curvature while the models with 3 stochastic factors predict negative coefficients for long maturities and positive coefficients for short maturities.

The evidence on conditional volatility is largely consistent with the results in Brandt and Chapman (2003) who conclude that quadratic term structure models provide a better fit to US term structure data than essentially affine models and point to conditional volatility regression results similar to the regression in Table 12 as the most important factor in the difference of fit between the two model classes[14].

[14]To test the robustness of the results a volatility regression is run in Appendix E where the dependent variable is yearly instead of monthly volatility and an ARCH term is added as an explanatory variable. The results are very similar.

Why do three-factor affine models fail to replicate the conditional regression coefficients? To provide a possible answer it is useful to compare model-implied conditional volatility with an estimate of actual conditional volatility. As a proxy for actual conditional volatility the conditional volatility from a EGARCH(1,1) model is estimated. Figure 4 and 5 graphs the model-implied and EGARCH(1,1) conditional volatility for the 1-year and 5-year yields[15].

[Figure 4 about here.]

[Figure 5 about here.]

The figures show that all models largely capture the persistence in conditional volatility apart from the period in the beginning of the 80's. While the EGARCH estimate is outside the confidence band for the models in some periods the trend is the same. It is also clear from the figures that both essentially and extended affine models fail to capture the high volatility during the Fed experiment from October 1979 to October 1982. Volatility in the 1-year yield across the models is about half the size of the EGARCH volatility and for the 5-year yield the model-implied volatility is roughly 50 % lower. To investigate the influence of the Fed experiment on the volatility regression results Table 13 shows the regression coefficients for the period before and after the Fed experiment.

[Table 13 about here.]

Compared to the coefficients obtained using the whole sample the results are quite different: the level coefficient is only about half the size before the Fed experiment and about one-third after the Fed experiment. In both subperiods the level effect is still highly significant. In addition, the slope coefficient is positive in both periods while it is negative for four out of five maturities when looking at the whole sample. The sign of the curvature coefficient is positive in the whole sample and both periods.

The positive slope and curvature coefficients are consistent with results in Christiansen and Lund (2005) (who argue that 1979-1982 should be excluded in analysis of the relation between the yield curve and volatility) and

[15] The model-implied conditional volatility is calculated for each of the 20,000 draws in the MCMC sampler and the mean and confidence band of the time $t = 1, ..., 631$ conditional volatility is estimated on basis of the 20,000 draws of time t conditional volatility.

the volatility regression results for the two subperiods are more in line with model-implied regression coefficients. The size of the level effect in the affine models are largely consistent with the size in the two subperiods. For example, the essentially affine $A_2(3)$ model has level coefficients that statistically match the actual level coefficients for all maturities for the period before the Fed experiment while the extended affine $A_1(3)$ model has level coefficients matching all the coefficients for the period after the Fed experiment. For the $A_1(3)$ and $A_2(3)$ models the positive slope and curvature coefficients are consistent with the positive coefficients found in the subperiods while the negative coefficients for the $A_3(3)$ models are not. The size of the positive slope and curvature coefficients in the $A_1(3)$ models are comparable with those estimated in the two subperiods while they are generally too high in the $A_2(3)$ models. Therefore, comparing the model-implied coefficients with the coefficients from the subperiods leads to the conclusion that $A_1(3)$ models capture volatility well and that the ability to capture volatility dynamics of yields does not improve with the number of stochastic volatility factors. The choice of risk premium does not change the conclusion. This might be because the advantage of increasing the number of factors entering volatility is outweighed by the more restricted correlation structure between factors.

Table 14 shows that the correlations between EGARCH and model-implied volatility are positive and in the range of 69.3% to 81.4% across models and maturity for the whole sample and similar correlations are found in the subperiods before and after the Fed experiment although the latter period has somewhat smaller correlations. This result is consistent with Almeida, Graveline, and Joslin (2006) who find similar positive correlations between conditional volatilities of yields of different maturities and GARCH estimates, while Collin-Dufresne, Goldstein, and Jones (2005) find that an extended $A_1(3)$ model generates a time series of volatility that is negatively correlated with a GARCH estimate of the short rate volatility. The differences might be due to different sample periods and that the short rate has quite different volatility dynamics from longer-maturity yields[16].

[Table 14 about here.]

Comparing essentially and extended models, the correlations suggest that extended $A_2(3)$ and $A_3(3)$ models do slightly better in terms of capturing conditional volatility than their essentially affine counterparts since they have

[16]See Joslin (2006) for an elaboration on this points.

higher average correlations in the whole period and all subsamples. Ahn, Dittmar, and Gallant (2002) find that completely affine models fare very poorly in capturing the conditional volatility of yield changes and evidence in Dai and Singleton (2003) suggests that persistence of volatility is larger in essentially affine models than in completely affine models. Therefore, the literature suggests that there are gains in terms of matching time-varying volatility in moving from completely affine to essentially affine models and the correlations suggest that gains in moving from essentially affine to extended affine models are positive but small for the $A_2(3)$ and $A_3(3)$ models. However, the results also show that there is no clear difference between the essentially and extended affine $A_1(3)$ models and the regression coefficients and correlations suggest that the $A_1(3)$ model does better than the $A_2(3)$ and $A_3(3)$ models. Therefore, there is no clear evidence showing that any of the extended models match volatility better than the essentially affine $A_1(3)$ model.

The comparison between model-implied conditional volatility and an EGARCH estimate is sensitive to possible misspecification of the EGARCH model and as an alternative procedure to compare the models I employ the use of realized volatility in testing volatility dynamics as suggested by Andersen and Benzoni (2006). They show that in affine models the quadratic variation of zero coupon yields is affine in average yields and they term this relation *the fundamental affine yield variation spanning condition*. Using intra-day data they approximate quadratic variation with realized volatility and test the spanning condition on daily data over the period 1991-2001 by regressing realized volatility on average yields. They find low R^2's and conclude that volatility factors are largely unrelated to the cross-section of yields.

To test the spanning condition for the longer time series in this paper I calculate a monthly realized volatility measure by summing squared daily returns over the month and regress it on the average level, slope, and curvature. The data are U.S. Treasury yield curve estimates of the Federal Reserve Board from July 1961 to the end of the sample[17]. Table 15 shows actual and model-implied R^2's for the period with available daily data 1961-2004 and the subperiod after the Fed experiment 1982-2004. The results for the sub-

[17]Gürkaynak, Sack, and Wright (2006) explain the estimation methodology by which the zero coupon yields are extracted and the data can be downloaded at http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html. The total number of daily observations is 10,834 for each maturity. Daily data are not available before 1961.

period before the Fed experiment are similar to the results for the whole sample period and are not shown.

[Table 15 about here.]

The results in the table show that for the whole sample period the R^2's between 31.1% and 34.5% are in line with those implied from the models: both $A_2(3)$ models and the extended $A_1(3)$ model have lower model-implied R^2's while the other three models have higher R^2's. Only the essentially affine $A_3(3)$ model has statistically significant different R^2's. Compared to the whole period 1961-2004 the subperiod 1982-2004 has lower R^2's ranging from 5.2% to 8.8%. All models have higher model-implied R^2's but the difference between actual and model-implied R^2's are insignificant for all but the $A_3(3)$ models. The spanning test therefore provides further evidence against the ability of the $A_3(3)$ models to model volatility while the $A_1(3)$ and $A_2(3)$ models pass the test.

Overall, the results in this section show that extended $A_2(3)$ and $A_3(3)$ models have slightly better volatility dynamics than their essentially affine counterparts, but volatility dynamics in $A_1(3)$ models - where the difference between extended and essentially affine is small - are most in accordance with moments in historical data. The evidence that an essentially or extended affine $A_1(3)$ model captures time-varying volatility well in a time span of more than 50 years apart from the 3 years during the Fed experiment suggests that if it is of importance to fit volatility correctly within a 3-factor affine framework in the whole period a regime-switching model might be needed. Results in Bansal, Tauchen, and Zhou (2004) support this conclusion. On the other hand, if it is not believed that an event such as the Fed experiment is likely to happen again an $A_1(3)$ essentially or extended affine model captures the volatility dynamics well.

5.4 Distribution of Yields: Extended Risk Premium and the Feller Condition

In the earlier sections we have seen how well essentially and extended affine models capture the first two moments of bond yields. In this section we will see how well the models capture higher-order moments. To see this, Figure 6 shows the unconditional distribution of the 5-year yield. The density of the actual 5-year yield is based on 631 historical observations while the densities of the model-implied 5-year yields are based on 10,000 years of simulated data taking the estimated parameters as the true parameters.

[Figure 6 about here.]

From the figure it is clear that the extended models capture the distribution better than the essentially affine models. The distribution in the essentially affine models is more skewed than the distribution of the actual data and the skew is stronger with fewer number of volatility factors. In contrast, the extended models do a reasonable job of capturing the distribution although the models put too little weight on low yields.

To help shed light on the reason why the distribution of the 5-year yield is so skewed in the essentially affine models it is helpful to compare the essentially affine and extended $A_1(3)$ models. The $A_1(3)$ models are chosen for comparison since they are the most commonly used 3-factor models with stochastic volatility, their distributions differ the most, and they provide the clearest intuition behind the difference.

Empirically, the volatility factor in $A_1(3)$ models is typically highly correlated with the yield with longest maturity - in this case the 5-year yield - and has a slow mean-reversion under the risk-neutral measure while the other two factors has a higher mean-reversion and often correspond to the slope and curvature of the yield curve. This is also the case in the $A_1(3)$ models estimated in this paper. In principle there might be more than one factor with a slow mean-reversion but this would limit the model's ability to fit a wide variety of term-structure shapes[18]. Since the two non-volatility factors 'die out' rather quickly the 5-year yield is close to being modelled as an affine function of a one-factor CIR process. As an approximation it is therefore reasonable to assume that the 5-year yield is an affine function of the volatility factor X_t,

$$\begin{aligned} Y_t^5 &\simeq \delta_0 + \delta_x(1)X_t \\ dX_t &= (K_0(1) - K_1(1,1)X_t)dt + \sqrt{X_t}dW_t. \end{aligned}$$

X_t is unconditionally gamma distributed with skewness $\sqrt{\frac{2}{K_0(1)}}$ and excess kurtosis $\frac{3}{K_0(1)}$ and since kurtosis and the absolute value of skewness are

[18]For a more detailed discussion on why one of the factors typically has a slow mean-reversion and the other two factors a higher mean-reversion see Duffee (2002).

unchanged by an affine transformation the 5-year yield Y^5 has skewness $sign(\delta_x(1))\sqrt{\frac{2}{K_0(1)}}$ and kurtosis $\frac{3}{K_0(1)}$. While all the parameters $\delta_0, \delta_x(1), K_0(1)$, and $K_1(1,1)$ determine the mean of X_t and three of the parameters determine the variance of X_t the only parameter determining kurtosis and the absolute value of skewness is $K_0(1)$. When the Q and P dynamics share parameters the Q dynamics usually dominate the P dynamics in terms of determining the parameters in estimation and therefore in the essentially affine $A_1(3)$ model the skewness and kurtosis of Y^5 under P is primarily determined by the skewness and kurtosis under Q since the Q and P dynamics share the parameter $K_0(1)$[19]. In contrast, the parameter $K_0(1)$ is allowed to differ under P and Q in the extended $A_1(3)$ model and this dramatically changes the skewness and kurtosis under P such that the model-implied 3. and 4. moments of the 5-year yield more closely resembles the historical 3. and 4. moments. To put it simple, the actual distribution of the 5-year yield in the essentially affine $A_1(3)$ model inherits a counterfactual skewed and fat-tailed distribution from the risk-neutral dynamics while the extra risk premium parameter in the extended affine $A_1(3)$ model allows the skewness and kurtosis of the distribution under P and Q to be different.

Although extended models have more flexibility in generating distributions with different skewness and kurtosis under P and Q, the models are more restricted in generating highly skewed and fat-tailed distributions. The Feller condition requires $K_0(1) > 0.5$ and therefore the skewness and excess kurtosis in the unconditional distribution of the 5-year yield cannot be higher than 2 and 1.5 in the $A_1(3)$ model. This restriction is even tighter when looking at the conditional distribution of $Y^5_{t+\tau|t}$ given the information at time t. As shown in Appendix F excess kurtosis and the absolute value of skewness of $Y^5_{t+\tau|t}$ are monotone increasing in τ, go to zero as τ goes to zero, and go to the excess kurtosis and the absolute value of skewness of the unconditional distribution as $\tau \to \infty$. While this restriction is not binding under P according to the estimates of the extended $A_1(3)$ model - $K_0^P(1)$ is estimated to be much larger than 0.5 - it appears to be binding under the risk-neutral measure since the estimate of $K_0^Q(1)$ is less than 0.5 in the

[19]An indication that the cross-section of yields seems to dominate the time-series properties of yields in terms of estimating parameters can be seen in the parameter estimates in the $A_1(3)$ models. When the parameter $K_0^P(1)$ is allowed to differ from $K_0^Q(1)$ in the extended model it is estimated at 2.2731 while in the essentially affine model where $K_0^P(1) = K_0^Q(1)$ it is estimated at the much smaller value 0.3741.

essentially affine $A_1(3)$ model. Apparently the data calls for a skewed and fat-tailed distribution for the 5-year yield under the risk-neutral measure and the extended models cannot accommodate this because of the Feller restriction. The limited ability of the extended $A_1(3)$ model to generate strongly skewed and fat-tailed distributions is illustrated in Figure 7.

[Figure 7 about here.]

In the figure, the conditional distribution under the risk-neutral measure of the 5-yield in December 2009 given information up to December 2004 - the end of the sample - is graphed[20]. It is seen that the essentially affine $A_1(3)$ model is more skewed than the extended $A_1(3)$ model and put more weight on high interest rates, while the smaller skew in the extended model has the effect of moving probability mass from the right tail to the left tail implying higher risk-neutral probabilities of small interest rates and smaller probabilities of high interest rates. As an example, the risk-neutral probability of a 5-year rate higher than 9% is 2.1% in the essentially affine model and 0.9% in the extended model while rates lower than 2% is 0.8% respectively 1.8%. While I have focused on the longest maturity in the sample for interpretation purposes, Figure 6 and 7 are very similar for shorter maturities suggesting that the conclusions in this section are relevant for all maturities.

This example indicates how the Feller condition limits the ability of the extended $A_1(3)$ model in generating skewed distributions for the 5-year yield under the risk-neutral measure and therefore the extended models might be less successful in valuing instruments that are sensitive to the tail-behavior of yields under the risk-neutral measure such as out-of-the-money options. As an example of such instruments, Gupta and Subrahmanyam (2005) stress that for accurate pricing of out-of-the-money caps and floors it is important for the chosen term structure model to be able to fit the skew in the underlying interest rate distribution.

6 Conclusion

This paper compares extended affine models with essentially affine models to see whether the risk premium specification in extended models is able to improve essentially affine models' ability to capture risk premia and volatility.

[20]The conditional distributions are calculated on basis of parameter and latent factors estimates in the last of the 20,000 simulations in each model.

Extended affine models do better capture the positive relation between the slope of the yield curve and excess returns and their risk premia match the historical risk premia better. The improvement increases with the number of stochastic volatility factors. However, none of the extended models with stochastic volatility match the $A_0(3)$ model.

With regards to volatility both extended and essentially affine models capture the historical dynamics quite well except in the period of the Fed experiment in the beginning of the 80s. The extended models capture volatility slightly better than essentially affine models. In both models, the ability to capture time-varying volatility decreases with the number of volatility factors and an $A_1(3)$ model - essentially or extended - captures best the historical volatility dynamics.

Finally, it is shown that extended models match the unconditional distribution of yields better than essentially affine models. However, an examination of the $A_1(3)$ model shows that due to the Feller condition the extended model cannot generate the skewness and kurtosis under the risk-neutral measure in long yields that the data calls for.

The results have important implication for the choice of affine model. If one is mostly interested in the P-dynamics of yields, for example for risk management purposes, one should choose an extended affine model since extended models have better P-dynamics. In contrast, if the focus is on the Q dynamics, for example in pricing out-of-the-money options, an essentially affine model is preferred because the Feller condition restricts the risk-neutral moments in yields.

A Adjusting the Campbell-Shiller Regressions for Time-varying risk premia

In Dai and Singleton (2002) it is shown how the regression coefficients of one can be restored in the Campbell-Shiller regressions by adjusting for time-varying risk premia. In this section I restate their results.

Rearranging the terms in the definition of the one-period excess return on an n-period bond, $D(t+1,n) = \ln(\frac{P(t+1,n-1)}{P(t,n)}) - r_t$ and using the definition of zero-coupon yields results in

$$Y(t+1,n-1) - Y(t,n) + \frac{1}{n-1}D(t+1,n) = \frac{1}{n-1}(Y(t,n) - r(t)). \quad (8)$$

Defining the *yield term premium* and the *forward term premium* as

$$\begin{aligned} c(t,n) &\equiv Y(t,n) - \frac{1}{n}\sum_{i=0}^{n-1} E_t[r(t+i)] \\ p(t,n) &\equiv f(t,n) - E_t[r(t+n)], \end{aligned}$$

where $f(t,n) \equiv -\ln \frac{P(t,n+1)}{P(t,n)}$ is the n-forward rate, the realized excess return can be decomposed in a premium part $D^*(t+1,n)$ and an expectations part[21],

$$D(t+1,n) = D^*(t+1,n) + \sum_{i=1}^{n-1}\Big[E_t[r(t+i)] - E_{t+1}[r(t+i)]\Big],$$

where

$$D^*(t+1,n) = -(n-1)[c(t+1,n-1) - c(t,n-1)] + p(t,n-1). \quad (9)$$

Since $E_t[D(t+1,n)] = E_t[D^*(t+1,n)]$ economic content to equation (8) is added by taking conditional expectation

$$E_t\Big[Y(t+1,n-1) - Y(t,n) + \frac{1}{n-1}D^*(t+1,n)\Big] = \frac{1}{n-1}(Y(t,n) - r(t)).$$

The regression coefficients in these "risk-premium adjusted" Campbell-Shiller regressions are calculated using historical yields and model implied risk premia.

[21] For detailed calculations see Dai and Singleton (2002).

B Three-Factor Affine Models

In this section I review the canonical representation of all 3-factor affine models. For all models the process X is restricted to be stationary under P which is ensured by restricting the real part of the eigenvalues of the mean-reversion matrix to be positive.

B.1 $A_0(3)$

The representation of the $A_0(3)$ model is

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = -\begin{bmatrix} K_1^Q(1,1) & 0 & 0 \\ K_1^Q(2,1) & K_1^Q(2,2) & 0 \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix}\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} dt + d\widetilde{W}(t).$$

The matrix K_1 is lower triangular to ensure identification. The essentially affine market price of risk is

$$S_t^{\frac{1}{2}}\Lambda_t = \begin{pmatrix} \lambda_1(1) + \lambda_2(1,1)X_t^1 + \lambda_2(1,2)X_t^2 + \lambda_2(1,3)X_t^3 \\ \lambda_1(2) + \lambda_2(2,1)X_t^1 + \lambda_2(2,2)X_t^2 + \lambda_2(2,3)X_t^3 \\ \lambda_1(3) + \lambda_2(3,1)X_t^1 + \lambda_2(3,2)X_t^2 + \lambda_2(3,3)X_t^3 \end{pmatrix}.$$

The extended affine market price of risk does not extend the flexibility of the essentially affine market price of risk. For the purpose of identification the vector δ_x in equation (2) has to be non-negative.

B.2 $A_1(3)$

The $A_1(3)$ has the representation

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \left(\begin{bmatrix} K_0^Q(1) \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} K_1^Q(1,1) & 0 & 0 \\ K_1^Q(2,1) & K_1^Q(2,2) & K_1^Q(2,3) \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix}\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}\right) dt + diag\left(\begin{bmatrix} \sqrt{X_t^1} \\ \sqrt{1+\beta_2(1)X_t^1} \\ \sqrt{1+\beta_3(1)X_t^1} \end{bmatrix}\right) d\widetilde{W}(t).$$

For the process to be well defined the restrictions $K_0^Q(1) > 0$, $\beta_2(1) > 0$, $\beta_3(1) > 0$, and $K_1^Q(1,1) > 0$ apply[22]. For identification the second and third

[22]This parameterization is used in Cheridito, Filipovic, and Kimmel (2006) and is a consequence of employing the *invariant affine transformation* $T_A X(t) = X(t) + (0, \begin{bmatrix} k_{22} & k_{23} \\ k_{32} & k_{33} \end{bmatrix}^{-1} \begin{pmatrix} k_{21} \\ k_{31} \end{pmatrix})'$ to the canonical $A_1(3)$ model in Dai and Singleton (2000). The transformation leaves all parameters unchanged except δ and θ. The condition $K_1^Q(1,1) > 0$ is due to the condition $[(K_1^Q)^{-1}K_0^Q]_1 > 0$.

element of δ_x in equation (2) has to be non-negative.

The extended affine market price of risk is given as

$$S_t^{\frac{1}{2}}\Lambda_t = \begin{pmatrix} \boxed{\lambda_1(1)} & + & \lambda_2(1,1)X_t^1 & & & & \\ \lambda_1(2) & + & \lambda_2(2,1)X_t^1 & + & \lambda_2(2,2)X_t^2 & + & \lambda_2(2,3)X_t^3 \\ \lambda_1(3) & + & \lambda_2(3,1)X_t^1 & + & \lambda_2(3,2)X_t^2 & + & \lambda_2(3,3)X_t^3 \end{pmatrix}. \quad (10)$$

For X to be well defined under P $\lambda_1(1)$ has to satisfy the constraint $\lambda_1(1) \geq \frac{1}{2} - K_0^Q(1)$.

The extended affine model allows $\lambda_1(1)$ to be non-zero in contrast to the essentially affine model[23]. Since the essentially affine model nests the completely affine, the extended affine model has a larger number of risk premium parameters than the completely affine. The cost of this flexibility is that the inequality $K_0^Q(1) > \frac{1}{2}$ has to be satisfied in contrast to the inequality $K_0^Q(1) > 0$ in both the essentially and completely affine model. Because of this constraint - which turns out to be binding when turning to the data - the extended affine model nests neither the essential nor the completely affine models.

B.3 $A_2(3)$

The representation of the $A_2(3)$ model is[24]

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \left(\begin{bmatrix} K_0^Q(1) \\ K_0^Q(2) \\ 0 \end{bmatrix} - \begin{bmatrix} K_1^Q(1,1) & K_1^Q(1,2) & 0 \\ K_1^Q(2,1) & K_1^Q(2,2) & 0 \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix}\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}\right)dt + diag\left(\begin{bmatrix} \sqrt{X_t^1} \\ \sqrt{X_t^2} \\ \sqrt{1+\beta_3(1)X_t^1+\beta_3(2)X_t^2} \end{bmatrix}\right)d\widetilde{W}(t).$$

with restrictions $K_0^Q(i) > 0, \beta_3(i) > 0, [(K_1^Q)^{-1}K_0^Q]_i > 0,\ i = 1,2,\ K_1^Q(2,1) \leq 0,\ K_1^Q(1,2) \leq 0$, and $\delta_x(3) > 0$.

The extended market price of risk is given as

$$S_t^{\frac{1}{2}}\Lambda_t = \begin{pmatrix} \boxed{\lambda_1(1)} & + & \lambda_2(1,1)X_t^1 & + & \boxed{\lambda_2(1,2)}X_t^2 & & \\ \boxed{\lambda_1(2)} & + & \boxed{\lambda_2(2,1)}X_t^1 & + & \lambda_2(2,2)X_t^2 & & \\ \lambda_1(3) & + & \lambda_2(3,1)X_t^1 & + & \lambda_2(3,2)X_t^2 & + & \lambda_2(3,3)X_t^3 \end{pmatrix}.$$

[23] $S_t\Phi_1 + I^-\Phi_2 X_t$ in equation (6) can be reparameterized as in equation (10) with $\lambda_1(1) = 0$. In the rest of the paper I will use the latter parametrization for the essentially affine models for easier comparison of risk premium parameter estimates.

[24] The affine transformation $T_A X(t) = X(t) + (0, 0, -\frac{K_0^Q(3)}{K_1^Q(3,3)})'$ is performed on the canonical $A_2(3)$ model of Dai and Singleton (2000).

The risk premium parameters are subject to the constraints $\lambda_1(i) \geq \frac{1}{2} - K_0^Q(i), [(K_1^P)^{-1}K_0^P]_i > 0,\ i = 1,2, \lambda_2(1,2) \geq K_1^Q(1,2)$, and $\lambda_2(2,1) \geq K_1^Q(1,2)$. The four boxed parameters are the extra parameters the extended affine model provides in comparison with the essentially affine model.

The added restrictions the extended affine model places on the Q-parameters in contrast to the essentially and completely affine models are $K_0^Q(i) > \frac{1}{2}, i = 1,2$.

B.4 $A_3(3)$

The representation of the $A_3(3)$ model is

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \left(\begin{bmatrix} K_0^Q(1) \\ K_0^Q(2) \\ K_0^Q(3) \end{bmatrix} - \begin{bmatrix} K_1^Q(1,1) & K_1^Q(1,2) & 0K_1^Q(1,3) \\ K_1^Q(2,1) & K_1^Q(2,2) & K_1^Q(2,3) \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} \right) dt + diag\left(\begin{bmatrix} \sqrt{X_t^1} \\ \sqrt{X_t^2} \\ \sqrt{X_t^3} \end{bmatrix} \right) d\widetilde{W}(t).$$

and restrictions for existence are $K_1^Q(i,j) \leq 0, i,j = 1,..,3, j \neq i, K_0^Q > 0$, and $(K_1^Q)^{-1}K_0^Q > 0$. The extended market price of risk is given as

$$S_t^{\frac{1}{2}}\Lambda_t = \begin{pmatrix} \boxed{\lambda_1(1)} + \lambda_2(1,1)X_t^1 + \boxed{\lambda_2(1,2)}X_t^2 + \boxed{\lambda_2(1,3)}X_t^3 \\ \boxed{\lambda_1(2)} + \boxed{\lambda_2(2,1)}X_t^1 + \lambda_2(2,2)X_t^2 + \boxed{\lambda_2(2,3)}X_t^3 \\ \boxed{\lambda_1(3)} + \boxed{\lambda_2(3,1)}X_t^1 + \boxed{\lambda_2(3,2)}X_t^2 + \lambda_2(3,3)X_t^3 \end{pmatrix}$$

subject to the constraints $\lambda_2(i,j) \leq K_1^Q(i,j), i,j = 1,..,3, j \neq i,\ \lambda_1 > \frac{1}{2} - K_0^Q$, and $(K_1^P)^{-1}K_0^P > 0$.

The extended affine model has the nine boxed parameters extra compared to the essential and completely affine models. The necessary extra conditions in the extended model are $K_0^Q > \frac{1}{2}$.

C Implementation Details

As explained in the text draws violating parameter constraints can simply be discarded according to Gelfand et al. (1992). However, in the extended affine model this procedure leads to practically rejecting every draw and therefore the RW-MH algorithm is used when sampling these parameters in the extended affine models[25].

[25] According to Gelfand, Smith, and Lee (1992) a risk premium element λ can be drawn conditional on the parameter constraint. For example, an element of λ_1 is restricted to

The efficiency of the RW-MH algorithm depends crucially on the variance of the proposal normal distribution. If the variance is too low, the Markov chain will accept nearly every draw and converge very slowly while it will reject a too high portion of the draws if the variance is too high. I therefore do an algorithm calibration and adjust the variance in the first million draws in the MCMC algorithm. Within each parameter block ($K_0^Q, K_1^Q, \beta, d, X$, and in the extended affine models λ_1 and λ_2) the variance of the individual parameters is the same, while across parameter blocks the variance may be different. Roberts, Gelman, and Gilks (1997) recommend acceptance rates close to $\frac{1}{4}$ for models of high dimension and therefore the standard deviation during the algorithm calibration is chosen as follows: Every 100'th draw the acceptance ratio of each block is evaluated. If it is less than 5 % the standard deviation is doubled while if it is more than 40 % it is cut in half. This step is prior to the burn-in period since the convergence results of RW-MH only applies if the variance is constant (otherwise the Markov property of the chain is lost). Despite this calibration the Markov chain has in some of the models not converged after an additional five million draws. In this case additional simulations in the burn-in period are carried out until there is no visible non-stationary behavior of the time-series of the univariate parameter draws.

The normal distribution of the risk premium parameters are found as follows. According to Bayes' theorem

$$\begin{aligned} p(\lambda|\Phi_{\backslash\lambda}, X, Y) &\propto p(Y|\Phi, X)p(\lambda|\Phi_{\backslash\lambda}, X) \\ &\propto p(X|\Phi)p(\lambda|\Phi_{\backslash\lambda}) \\ &\propto \prod_{i=1}^{N}\Big(\exp\Big(-\frac{1}{2\Delta_t}\sum_{t=1}^{T}\frac{[\Delta X_t - \mu_{t-1}^P\Delta_t]_i^2}{[S_{t-1}]_{ii}}]\Big) \end{aligned}$$

In the last line it is used that the priors are assumed to be independent and proportional to a constant such that the data dominate the results. Furthermore

$$\begin{aligned} \mu_t^P &= K_0^Q - K_1^Q X_t + \sqrt{S_t}\Lambda_t \\ &= K_0^Q - K_1^Q X_t + \lambda_1 + \lambda_2 X_t, \end{aligned}$$

$[a;\infty)$ due to the multivariate Feller condition. Denoting F as the unconditional distribution function of λ and drawing a uniform random variable U, λ can be drawn as $\lambda = F^{-1}[F(a) + U(1 - F(a))]$. However, this procedure is not computationally feasible since the constrained interval lies far in the tail of the unconditional distribution and therefore $F(a)$ cannot reliably be computed.

so in the expression $[\Delta X_t - \mu^P_{t-1}\Delta_t]_i$ all the individual elements λ^{ind} in the vector λ_1 and matrix λ_2 can be written as $a_t\lambda^{ind} - b_t$,

$$p(\lambda^{ind}|\Phi_{\backslash\lambda^{ind}}, X, Y) \quad \propto \quad \exp\Big(- \frac{1}{2\Delta_t}\sum_{t=1}^{T}\frac{(a_t\lambda^{ind} - b_t)^2}{[S_{t-1}]_{ii}}]\Big).$$

Using the result in Frühwirth-Schnatter and Geyer (1998) p. 10 I have that $p(\lambda^{ind}|\Phi_{\backslash\lambda^{ind}}, X, Y)$ is a normal distribution with

$$\begin{aligned} E(\lambda^{ind}) &= Qm \\ var(\lambda^{ind}) &= Q \end{aligned}$$

where

$$\begin{aligned} m &= \sum_{t=1}^{T}\frac{a_t b_t}{\Delta t[S_t]_{ii}} \\ Q^{-1} &= \sum_{t=1}^{T}\frac{a_t^2}{\Delta t[S_t]_{ii}}. \end{aligned}$$

The conditional of X_t depends only on neighboring X's and the sampling of the latent process X can for computational speed be done in two steps. First $X_0, X_2, ...$ are sampled and second $X_1, X_3, ...$ are sampled. Of the total computing time, solving the ODEs (4)-(5) takes up 70-80% of the computing time.

D Model-Implied Campbell-Shiller Regression Coefficients

In the text the coefficent ϕ_n from the Campbell-Shiller regression

$$Y(t+1, n-1) - Y(t,n) = const + \phi_n[\frac{Y(t,n) - Y(t,1)}{n-1}] + res$$

is calculated. The model-implied coefficient ϕ_n is calculated as follows[26]. First

$$\frac{cov(Y_{t+1}(n-1) - Y_t(n), \frac{Y_t(n)-Y_t(1)}{n-1})}{var(\frac{Y_t(n)-Y_t(1)}{n-1})} = \frac{(n-1)cov(B(n-1)X_{t+1} - B(n)X_t, X_t)[B(n) - B(1)]'}{[B(n) - B(1)]var(X_t)[B(n) - B(1)]'}.$$

[26] The error term in the measurement equation is ignored since this is a statistical tool and not part of the equilibrium model underlying the term structure model.

The well known formulas $EZ = E(E(Z|X_t))$ and $EZX_t = E(E(Z|X_t)X_t)$ yield

$$\Psi(X_t) \equiv E[B(n-1)X_{t+1} - B(n)X_t|X_t] = B(n-1)[(I - e^{-K\tau})\theta + e^{-K\tau}X_t] - B(n)X_t$$

where it is used that $E(X_{t+\tau}|X_t) = (I - e^{-K\tau})\theta + e^{-K\tau}X_t$. Furthermore

$$\begin{aligned} cov(B(n-1)X_{t+1} - B(n)X_t, X_t) &= E[\Psi(X_t)X_t] - E[\Psi(X_t)]E[X_t] \\ &= [B(n-1)e^{-K\tau} - B(n)]var(X_t), \end{aligned}$$

and $var(X_T) = lim_{t\to-\infty} var(X_T|X_t)$ where analytic formulas for $var(X_T|X_t)$ is given in the Appendix in Duffee (2002). The formulas in Duffee (2002) require the mean-reversion matrix K under the actual measure P to be diagonalizable. Occasionally this requirement is not satisfied and in this case the variance is calculated using the results in Fisher and Gilles (1996). Specifically, the expression in equation (3.8) in Fisher and Gilles (1996) is numerically integrated to obtain the desired result.

E Yearly Volatility Regression

To check the robustness of the volatility regression results in the text an alternative regression using the residuals from the Campbell-Shiller regressions is examined in this Appendix.

The squared residuals from the CS regressions in equation (1),

$$[E_t(Y(t+1,n-1) - Y(t,n)) - (Y(t+1,n-1) - Y(t,n)]^2 = [Y(t+1,n-1) - E_t(Y(t+1,n-1))]^2,$$

regressed on the level, slope, and curvature of the yield curve and an ARCH term yields the regression

$$\begin{aligned} [Y(t+1,n) - E_t(Y(t+1,n))]^2 = \quad & const + \\ & \phi_{n+1}(1)[Y(t,5)] + \\ & \phi_{n+1}(2)[Y(t,5) - Y(t,1)] + \\ & \phi_{n+1}(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] + \\ & \phi_{n+1}(4)[Y(t,n) - E_{t-1}(Y(t,n))]^2 + \\ & res. \end{aligned}$$

Regression results for the data are given in Table 16. The level and ARCH factors are strongly significant while the slope and curvature factors are insignificant. Table 17 and 18 shows results for the affine models using simulated data as explained in the text. The results are nearly identical to the results in the text: a) the $A_0(3)$ estimates the coefficients to 0, b) the models get the correct sign of the level coefficient but only roughly half the actual magnitude and the differences are statistically significant, and c) the $A_1(3)$ and $A_2(3)$ models imply positive slope and curvature coefficients while the $A_3(3)$ models imply negative coefficients.

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

F Conditional Moments of the CIR Process

According to Cox, Ingersoll, and Ross (1985) the CIR process

$$dX = k(\theta - X)dt + \sigma\sqrt{X}dW,$$

has a density of the conditional distribution of $X_{t+\tau}|X_t$ given by

$$f(X_{t+\tau}|X_t) = ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q(2(uv)^{\frac{1}{2}}),$$

where

$$\begin{aligned} c &= \frac{2k}{\sigma^2(1-e^{-k\tau})}, \\ u &= cX_t e^{-k\tau}, \\ v &= cX_{t+\tau}, \\ q &= \frac{2k\theta}{\sigma^2} - 1, \end{aligned}$$

and $I_q(\cdot)$ is the modified bessel function of the first kind of order q. It is seen that $2v$ has a non-central χ^2 distribution with $f = \frac{4k\theta}{\sigma^2}$ degrees of freedom and non-centrality parameter $\lambda = 2u$. The mean, variance, skewness, and excess kurtosis are

$$\begin{aligned}
E(X_{t+\tau}) &= \frac{1}{2c}(f+\lambda) = \theta(1-e^{-k\tau}) + X_t e^{-k\tau} \\
V(X_{t+\tau}) &= \frac{1}{4c^2}(2f+4\lambda) = \frac{\sigma^2(1-e^{-k\tau})^2}{2k}\theta + \frac{\sigma^2 e^{-k\tau}(1-e^{-k\tau})}{k} X_t \\
skew(X_{t+\tau}) &= \frac{2^{\frac{3}{2}}(f+3\lambda)}{(f+2\lambda)^{\frac{3}{2}}} = 2^{\frac{3}{2}} \frac{\frac{4k\theta}{\sigma^2} + 6cX_t e^{-k\tau}}{(\frac{4k\theta}{\sigma^2} + 4cX_t e^{-k\tau})^{\frac{3}{2}}} = \frac{1}{\sqrt{k}\sigma} \frac{4\theta + \frac{12}{1-e^{-k\tau}} X_t e^{-k\tau}}{(2\theta + \frac{4}{1-e^{-k\tau}} X_t e^{-k\tau})^{\frac{3}{2}}} \\
&= \frac{\sqrt{2}}{\sqrt{k\theta}\sigma} \frac{2\sqrt{2} + \frac{3}{\sqrt{2}}K}{(2+K)^{\frac{3}{2}}} \\
exkurt(X_{t+\tau}) &= \frac{12(f+4\lambda)}{(f+2\lambda)^2} = 12 \frac{\frac{4k\theta}{\sigma^2} + 8cX_t e^{-k\tau}}{(\frac{4k\theta}{\sigma^2} + 4cX_t e^{-k\tau})^2} = \frac{12}{k\sigma^2} \frac{4\theta + \frac{16}{1-e^{-k\tau}} X_t e^{-k\tau}}{(4\theta + \frac{8}{1-e^{-k\tau}} X_t e^{-k\tau})^2} \\
&= \frac{3}{k\theta\sigma^2} \frac{4+4K}{(2+K)^2}
\end{aligned}$$

where

$$K = \frac{4}{e^{k\tau}-1} \frac{X_t}{\theta}.$$

It is easily seen that skewness and excess kurtosis are monotone decreasing in K so they are monotone increasing in τ. As $\tau \to 0$ skewness and excess kurtosis go to zero and as $\tau \to \infty$ they go to the skewness and excess kurtosis of the unconditional distribution of X - $\sqrt{\frac{2}{k\theta\sigma^2}}$ and $\frac{3}{k\theta\sigma^2}$.

References

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002). Quadratic Term Structure Models: Theory and Evidence. *Review of Financial Studies 15*(1), 243–288.

Almeida, C., J. J. Graveline, and S. Joslin (2006). Do Options Contain Information About Excess Bond Returns. *Working Paper*.

Andersen, T. G. and L. Benzoni (2006). Can Bonds hedge Volatility Risk in the U.S. Treasury market? A Specification Test for Affine Term Structure Models. *Working paper*.

Bansal, R., G. Tauchen, and H. Zhou (2004). Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle. *Journal of Business and Economic Statistics 22*(4), 396–409.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Association Series B 36*, 192–236.

Bester, C. A. (2004). Random Field and Affine Models for Interest Rates: An Empirical Comparison. *Working Paper*.

Brandt, M. W. and D. A. Chapman (2003). Comparing Multifactor Models of the Term Structure. *Working Paper*.

Campbell, J. Y. and R. J. Shiller (1991). Yield Spreads and Interest Rate Movements: A Bird's Eye View. *Review of Economic Studies 58*, 495–514.

Chapman, D. A. and N. D. Pearson (2001). Recent Advances in Estimating Term-Structure Models. *Financial Analysts Journal 57*, 77–95.

Cheridito, P., D. Filipovic, and R. L. Kimmel (2006). Market Price of Risk Specifications for Affine Models: Theory and Evidence. *Forthcoming, Journal of Financial Economics*.

Christiansen, C. and J. Lund (2005). Revisiting the Shape of the Yield Curve: The Effect of Interest Rate Volatility. *Working Paper*.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2005). Can Interest Rate Volatility be Extracted From the Yield Curve? A Test of Unspanned Stochastic Volatility. *Working Paper*.

Cox, J., J. Ingersoll, and S. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica 53*, 385–407.

Dai, Q. and K. J. Singleton (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance 55*(5), 1943–1978.

Dai, Q. and K. J. Singleton (2002). Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure. *Journal of Financial Economics 563*, 415–441.

Dai, Q. and K. J. Singleton (2003). Term Structure Dynamics in Theory and Reality. *Review of Financial Studies 16*(3), 631–678.

Duffee, G. R. (2002). Term premia and Interest Rate Forecasts in Affine Models. *Journal of Finance 57*, 405–443.

Duffie, D. and R. Kan (1996). A Yield-factor Model of Interest Rates. *Mathematical Finance 6*, 379–406.

Elton, E. J. and T. C. Green (1998). Tax and Liquidity Effects in Pricing Government Bonds. *Journal of Finance 53*(5), 1533–1562.

Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics 19*, 177–191.

Fama, E. F. (1984). The Information in the Term Structure. *Journal of Financial Economics 13*, 509–528.

Fama, E. F. and R. Bliss (1987). The Information in Long-Maturity Forward Rates. *American Economic Review 77*, 680–692.

Fisher, M. and C. Gilles (1996). Term Premia in Exponential-Affine Models of the Term Structure. *Working Paper*.

Frühwirth-Schnatter, S. and A. L. Geyer (1998). Bayesian Estimation of Econometric Multifactor Cox Ingersoll Ross Models of the Term Structure of Interest Rates via MCMC Methods. *Working Paper, Department of Statistics, Vienna University of Economics and Business Administration*.

Gelfand, A. E., A. F. Smith, and T.-M. Lee (1992). Bayesian Analysis of Constrained Parameter and Truncated Data problems Using Gibbs Sampling. *Journal of the American Statistical Association 87*(418), 523–532.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 6*, 721–741.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2006). The U.S. Treasury Yield Curve: 1961 to the Present. *Working Paper, Divisions of Reasearch and Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.*.

Gupta, A. and M. G. Subrahmanyam (2005). Pricing and Hedging Interest Rate Options: Evidence from Cap-Floor Markets. *Journal of Banking and Finance 29*, 701–733.

Hammersley, J. and P. Clifford (1970). Markov Fields on Finite Graphs and Lattices. *Unpublished Manuscript*.

Hansen, L. P. and R. J. Hodrick (1980). Forward Exchange Rates as Optimal predictors of future Spot Rates: An Econometric Analysis. *Journal of Political Economy 88*(51), 829–853.

Johannes, M. and N. Polson (2003). MCMC Methods for Financial Econometrics. *forthcoming, Handbook of Financial Econometrics*.

Joslin, S. (2006). Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities? *Working Paper*.

Lamoureux, C. G. and H. D. Witte (2002). Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll, and Ross. *Journal of Finance 57*, 1479–1520.

Litterman, R. and J. Scheinkman (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income 1*, 54–61.

Marsaglia, G. and A. Zaman (1991). A New Class of Random Number Generators. *Annals of Applied Probability 3*, 462–480.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equations of State Calculations by fast Computing Machines. *Journal of Chemical Physics 21*, 1087–1091.

Mikkelsen, P. (2001). MCMC Based Estimation of Term Structure Models. *Working Paper*.

Piazzesi, M. (2003). Affine Term Structure Models. *forthcoming, Handbook of Financial Econometrics*.

Piazzesi, M. (2005). Bond Yields and the Federal Reserve. *Journal of Political Economy 113*(2), 311–344.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2 ed.). Springer-Verlag.

Roberts, G., A. Gelman, and W. Gilks (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Annals of Applied Probability 7*(1), 110–120.

Sanford, A. D. and G. M. Martin (2005). Simulation-Based Bayesian Estimation of an Affine Term Structure Model. *Computational Statistics and Data Analysis 49*, 527–554.

Tanner, M. and W. Wong (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association 82*, 528–549.

Figures

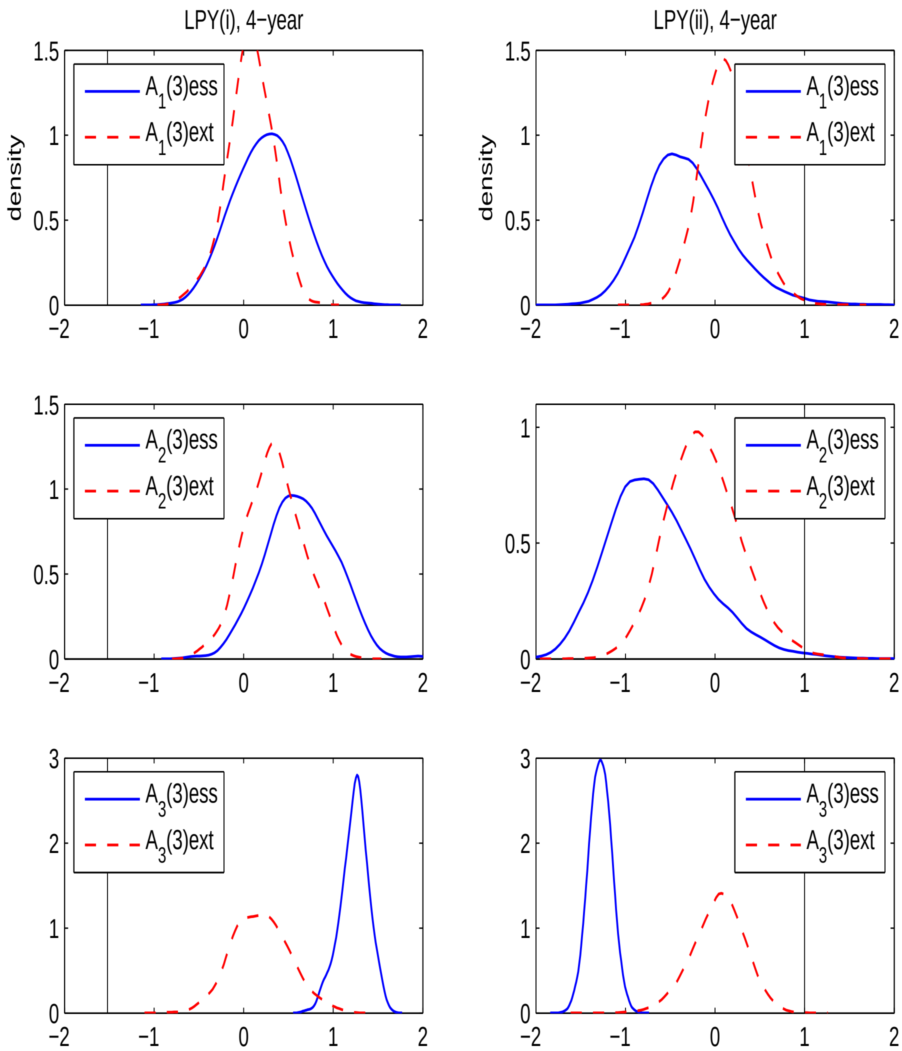

Figure 1: Density of unadjusted and adjusted Campbell-Shiller regression coefficients. The first column of this figure shows for each affine model the density of the Campbell-Shiller regression coefficient estimated according to equation (1). The coefficients are for the 4-year excess return (n=4) and the actual coefficient estimated from the data is -1.520 and is marked with a vertical solid line. The second column of this figure shows the adjusted Campbell-Shiller regression coefficient estimated using actual yields and model-implied risk premia. The coefficients are for the 4-year excess return (n=4) and the coefficient of 1 which the models should match if they are capturing time-varying risk premia correctly is marked with a vertical solid line.

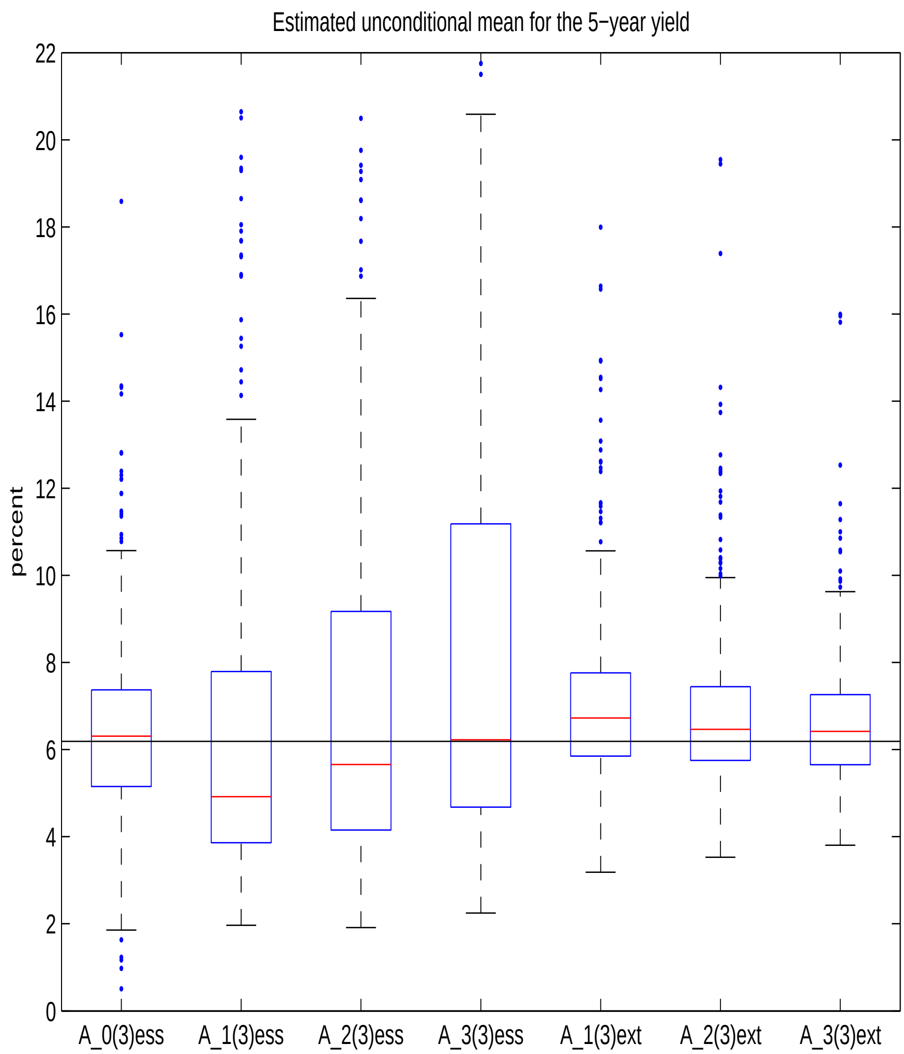

Figure 2: Box-plot of unconditional mean. This figure depicts for each affine model a box-plot of the density of the unconditional 5-year yield. The black horizontal line is the mean in the 5-year yield of 6.19 percent.

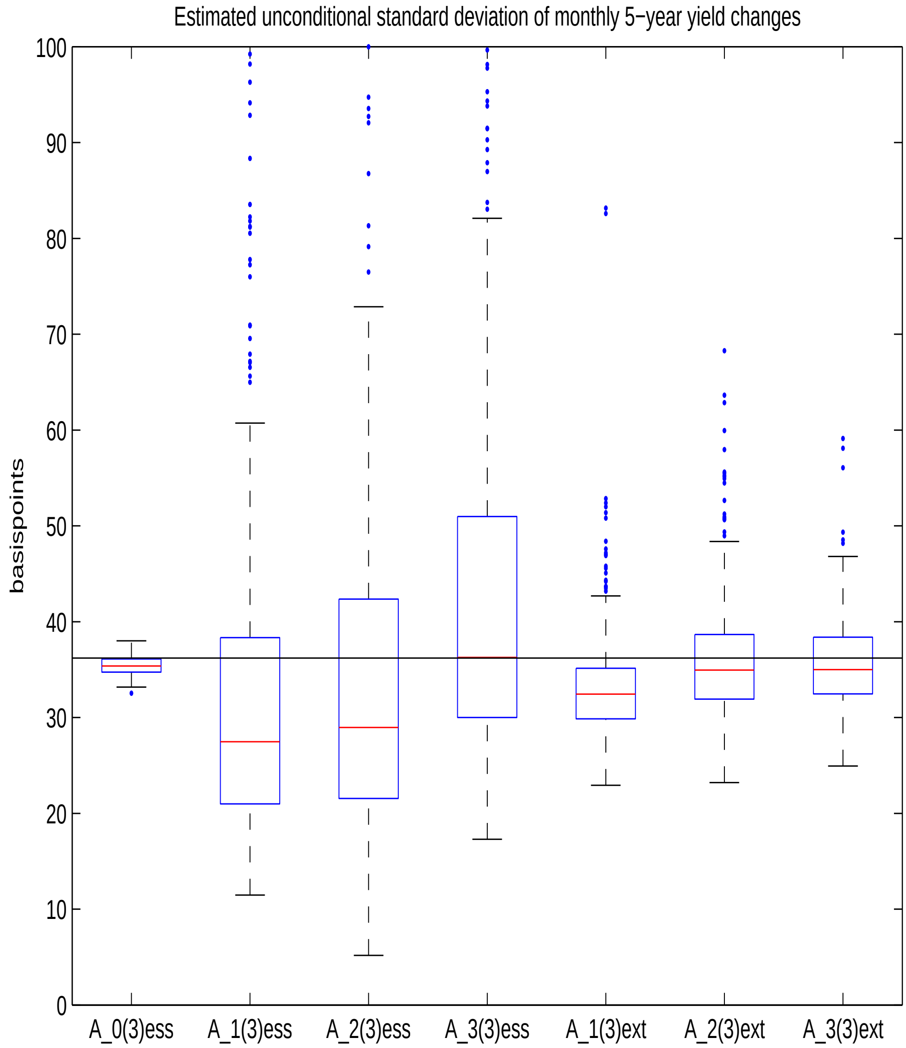

Figure 3: Box-plot of unconditional volatility. This figure depicts for each affine model a box-plot of the density of the unconditional 5-year volatility (standard deviation). The black horizontal line is the unconditional volatility in the 5-year yield of 36.2 basis points.

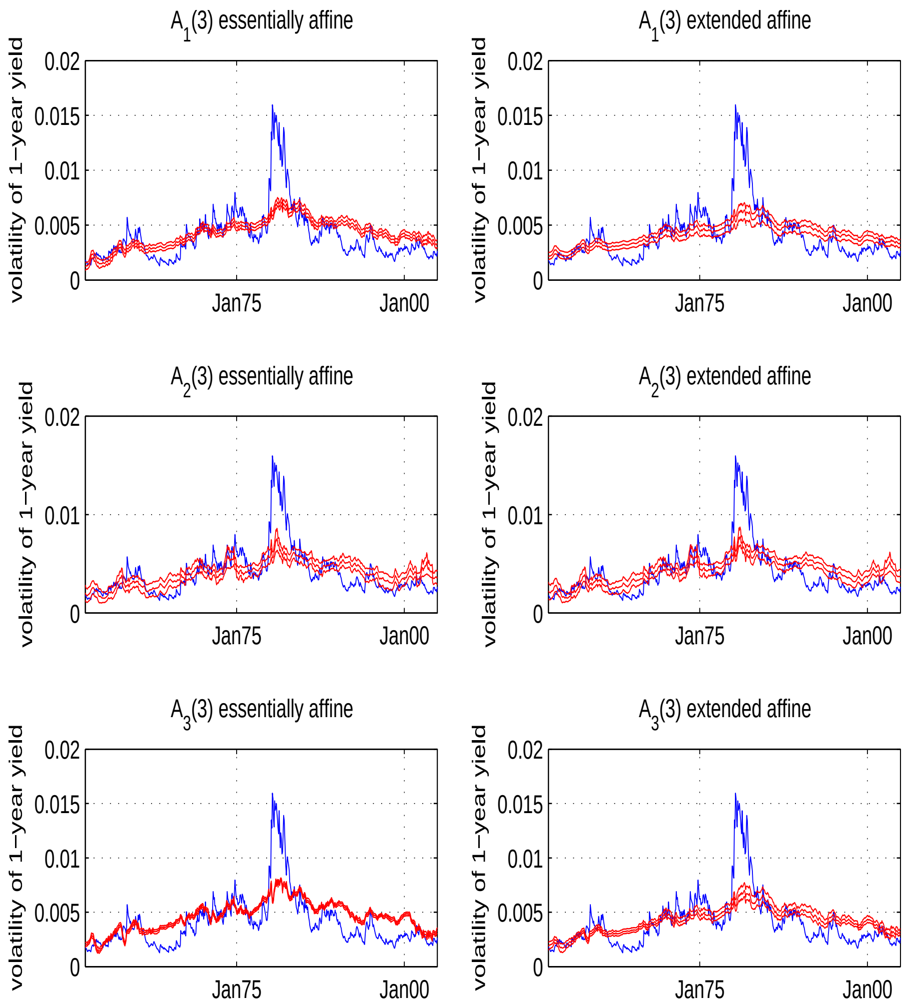

Figure 4: Actual and model-implied 1-year conditional volatility. This figure depicts for the affine models with stochastic volatility the estimate and confidence band of the monthly conditional volatility in the 1-year yield along with an EGARCH(1,1) estimate.

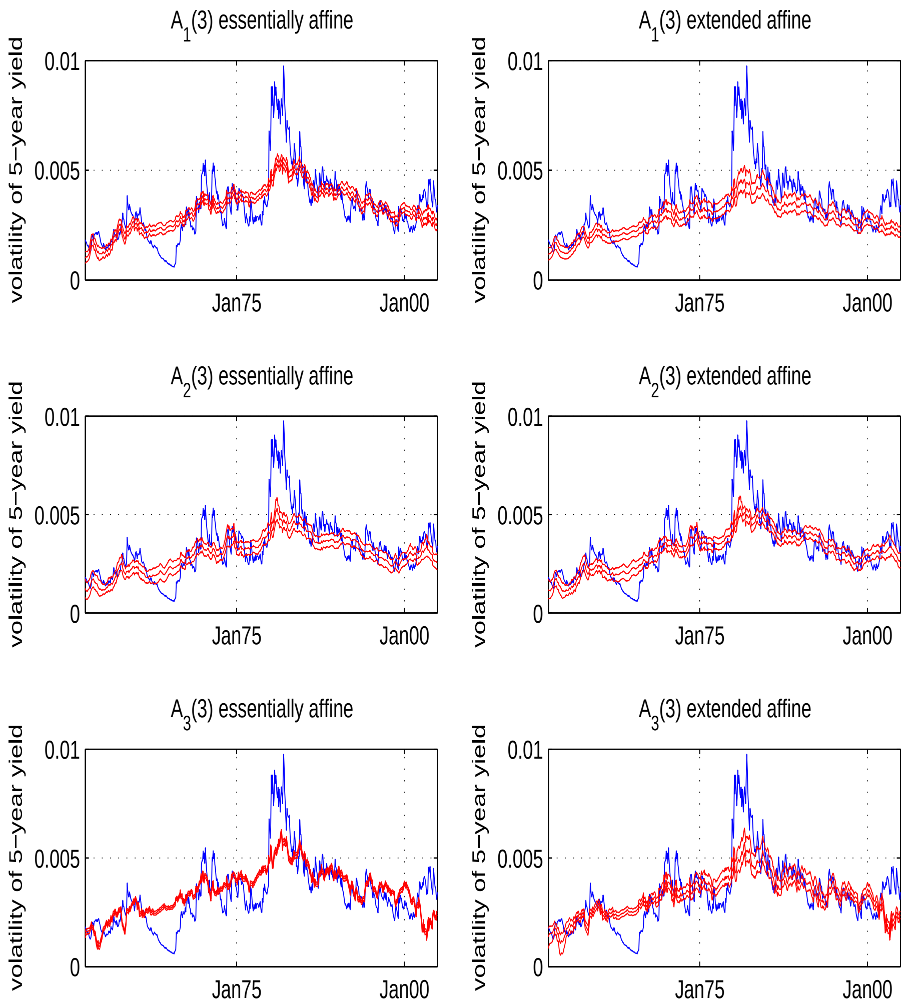

Figure 5: Actual and model-implied 5-year conditional volatility. This figure depicts for the affine models with stochastic volatility the estimate and confidence band of the monthly conditional volatility in the 5-year yield along with an EGARCH(1,1) estimate.

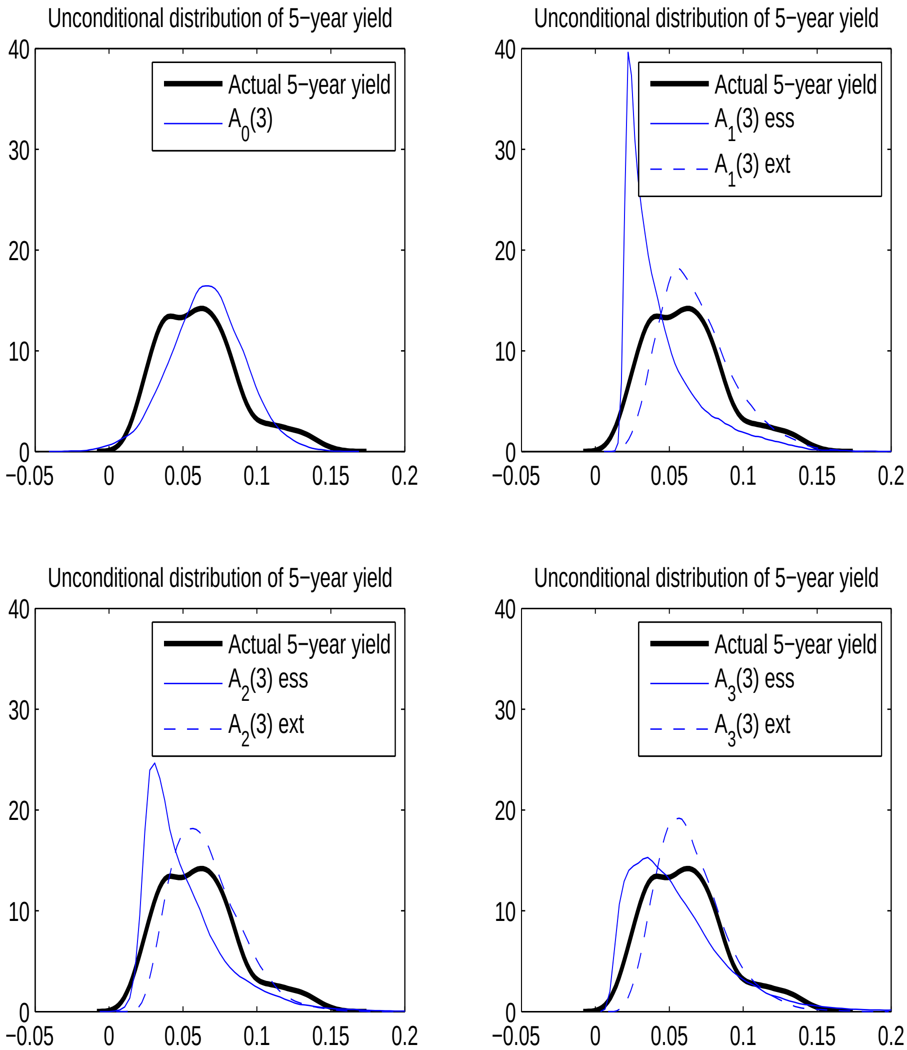

Figure 6: Actual and model-implied distribution of the 5-year yield. This figure shows for the actual 5-year yield and the 7 models the unconditional distribution of the 5-year yield.

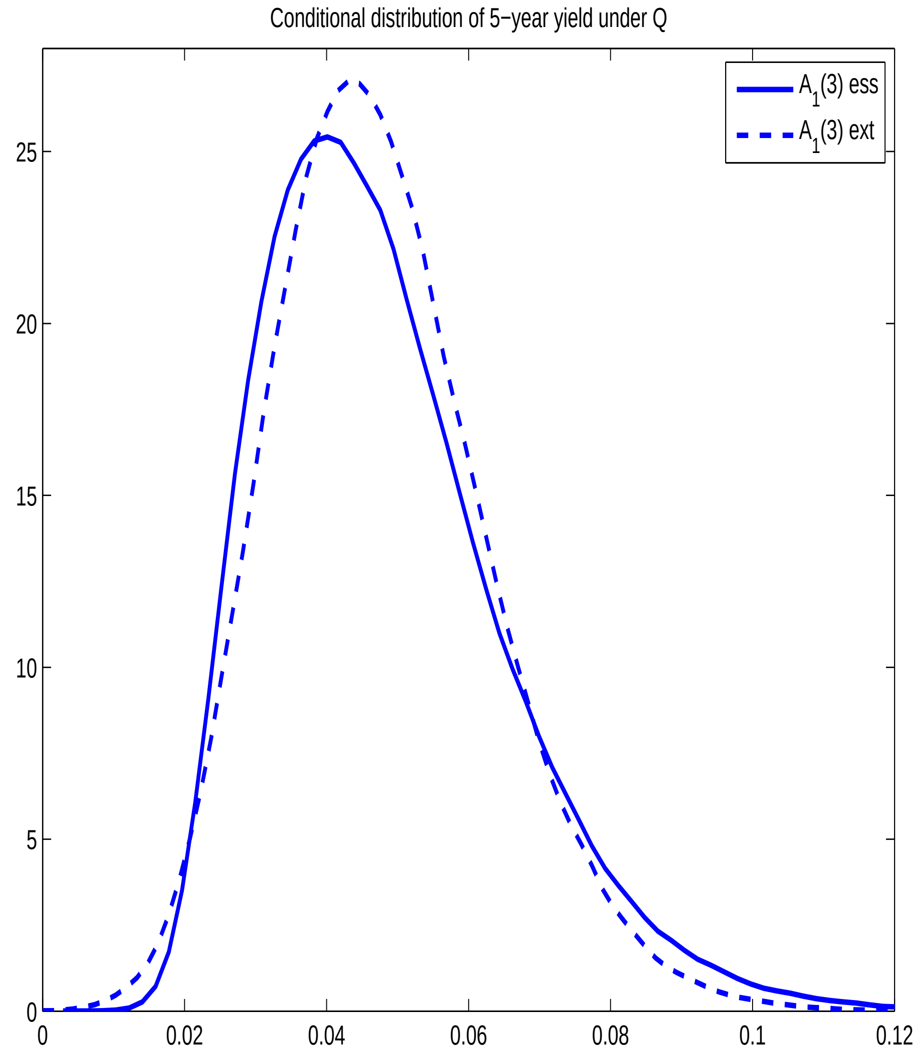

Figure 7: Conditional distribution of 5-year yield under the risk-neutral measure. This figure shows for the last observation in the sample, December 2004, the conditional distribution of the 5-year yield at December 2009 under the risk-neutral measure. The skewness and excess kurtosis for the distribution are 0.92 and 1.23 in the essentially affine $A_1(3)$ model while they are 0.51 and 0.47 in the extended $A_1(3)$ model.

Tables

n	1	2	3	4	5
mean	5.60	5.81	5.98	6.11	6.19
vol. of yield changes	49.3	43.2	40.1	38.8	36.2

Table 1: *Yield Curve Statistics.* This table shows the unconditional mean of yields along with the volatility (standard deviation) of monthly yield changes for maturities 1, 2, 3, 4, and 5 year. The mean is given in percent while the volatility is measured in basis points. Data: Fama and Bliss (1987) monthly observations from 1952:6 to 2004:12.

n	2	3	4	5
ϕ_n	-0.775^{**}	-1.1311^{***}	-1.5198^{***}	-1.4941^{***}
s.e.	(0.546)	(0.637)	(0.683)	(0.745)

Table 2: *Campbell-Shiller regressions.* This table shows the slope coefficients from the regressions $Y(t+1,n-1) - Y(t,n) = const + \phi_n[\frac{Y(t,n)-Y(t,1)}{n-1}] + residual$ where n and t are measured in years. In parenthesis are shown Hansen and Hodrick (1980) standard errors and a significant difference from 1 at the 5%, 1%, or 0.1% level is denoted by $*, **$, or $* * *$. Data: Fama and Bliss (1987) monthly observations from 1952:6 to 2004:12.

n	1	2	3	4	5
level	0.1095***	0.0713***	0.0565***	0.0519***	0.0438***
	(0.0202)	(0.0132)	(0.0102)	(0.0075)	(0.0058)
slope	−0.1415	−0.0785	−0.0316	−0.0196	0.0156
	(0.0960)	(0.0631)	(0.0479)	(0.0362)	(0.0286)
curvature	0.2712	0.1082	0.1657	0.0776	0.1262*
	(0.1968)	(0.1304)	(0.0974)	(0.0755)	(0.0603)

Table 3: *Volatility regression.* This table shows the coefficents from the regressions $[Y(t+1,n)-Y(t,n)]^2 = const+\phi_n(1)[Y(t,5)]+\phi_n(2)[Y(t,5)-Y(t,1)]+\phi_n(3)[Y(t,5)+Y(t,1)-2Y(t,3)] + residual$ where t is measured in months, n in years, and Y in percent. In parenthesis are shown Hansen and Hodrick (1980) standard errors with 6 lags and significance at the 5%, 1%, or 0.1% level is denoted by $*, **$, or $***$. Data: Fama and Bliss (1987) monthly observations from 1952:6 to 2004:12.

	$A_0(3)$	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess
$K_1^Q(1,1)$	0.6250 $(0.5344; 0.7075)$	0.0318 $(0.0100; 0.0550)$	1.1323 $(0.9843; 1.2964)$	0.2357 $(0.1987; 0.2850)$
$K_1^Q(1,2)$	0	0	-0.0770 $(-0.1439; -0.0182)$	-0.0461 $(-0.1067; -0.0026)$
$K_1^Q(1,3)$	0	0	0	-0.4445 $(-0.6668; -0.2844)$
$K_1^Q(2,1)$	4.6914 $(4.0994; 5.2303)$	3.5617 $(3.1709; 3.9545)$	-0.1634 $(-0.2978; -0.0301)$	-0.1661 $(-0.1884; -0.1385)$
$K_1^Q(2,2)$	8.6864 $(8.3999; 9.1701)$	0.0982 $(0.0107; 0.1700)$	0.0549 $(0.0222; 0.0978)$	0.1590 $(0.1355; 0.1894)$
$K_1^Q(2,3)$	0	4.0489 $(3.5131; 4.5243)$	0	-0.0781 $(-0.1988; -0.0011)$
$K_1^Q(3,1)$	1.5609 $(0.9862; 1.8184)$	1.9465 $(1.7062; 2.3039)$	2.4236 $(2.1944; 2.7326)$	-0.0033 $(-0.0117; -0.0001)$
$K_1^Q(3,2)$	2.8678 $(2.2982; 3.3807)$	-0.0735 $(-0.1056; -0.0459)$	-0.2465 $(-0.4159; -0.0620)$	-0.2937 $(-0.3181; -0.2667)$
$K_1^Q(3,3)$	0.0163 $(-0.0002; 0.0355)$	1.0179 $(0.8830; 1.1557)$	0.3566 $(0.2945; 0.4368)$	1.9138 $(1.8115; 1.9833)$
$K_0^Q(1)$	0	0.3741 $(0.1742; 0.5917)$	0.2169 $(0.0082; 0.6824)$	0.7475 $(0.3025; 1.2112)$
$K_0^Q(2)$	0	0	0.5095 $(0.0579; 1.4771)$	0.0764 $(0.0049; 0.1831)$
$K_0^Q(3)$	0	0	0	0.0427 $(0.0011; 0.1596)$
$\lambda_2(1,1)$	-0.0580 $(-0.4829; 0.3163)$	0.0005 $(-0.0551; 0.0413)$	0.1185 $(-0.1091; 0.3513)$	0.0998 $(0.0300; 0.1652)$
$\lambda_2(1,2)$	-0.2414 $(-0.7987; 0.2982)$	0	0	0
$\lambda_2(1,3)$	0.0237 $(-0.0962; 0.1429)$	0	0	0
$\lambda_2(2,1)$	4.1679 $(3.5749; 4.7157)$	1.7675 $(-4.8665; 8.4713)$	0	0
$\lambda_2(2,2)$	6.9890 $(6.2871; 7.7656)$	-0.2431 $(-0.5787; 0.0709)$	0.0208 $(-0.0247; 0.0634)$	-0.0838 $(-0.1626; -0.0093)$
$\lambda_2(2,3)$	-0.1225 $(-0.2451; -0.0023)$	3.0486 $(0.4158; 5.7882)$	0	0
$\lambda_2(3,1)$	1.2327 $(0.6157; 1.7127)$	-1.3007 $(-2.6867; -0.0567)$	0.9776 $(-0.7547; 2.7431)$	0
$\lambda_2(3,2)$	2.9864 $(2.1810; 3.7600)$	0.0665 $(0.0030; 0.1378)$	-0.1865 $(-0.4248; -0.0308)$	0
$\lambda_2(3,3)$	-0.0781 $(-0.1976; 0.0319)$	-0.5906 $(-1.1165; -0.1028)$	-0.0267 $(-0.2714; 0.2161)$	0.0251 $(-0.2209; 0.2709)$
σ^2	$4.90e-7$ $(4.57; 5.24)e-7$	$4.93e-7$ $(4.61; 5.28)e-7$	$5.21e-7$ $(4.86; 5.59)e-7$	$5.11e-7$ $(4.78; 5.48)e-7$

Table 4: Model estimates, essentially affine models (part 1). This Table shows parameter estimates along with confidence bands for all 3-factor essentially affine models as defined in section B. Data used in estimation is Fama and Bliss (1987) monthly zero yields from 1952:6 to 2004:12. The estimation method is MCMC.

	$A_0(3)$ ess/ext	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess
$\lambda_1(1)$	0.5289 (0.1485; 0.9139)	0	0	0
$\lambda_1(2)$	−0.1127 (−0.5192; 0.2938)	0.3850 (−56.14; 55.90)	0	0
$\lambda_1(3)$	0.2210 (−0.1697; 0.6163)	−0.2580 (−10.90; 10.74)	0.2958 (−2.0638; 2.7795)	0
δ_0	0.0790 (0.0762; 0.0826)	0.0187 (0.0184; 0.0191)	0.0227 (0.0218; 0.0236)	0.0112 (0.0100; 0.0124)
$\delta_x(1)$	0.0192 (0.0134; 0.0282)	0.0027 (0.0025; 0.0029)	0.0071 (0.0052; 0.0089)	−0.0014 (−0.0015; −0.0012)
$\delta_x(2)$	0.0684 (0.0637; 0.0750)	0.00006 (0.00004; 0.00008)	0.0007 (0.0006; 0.0010)	0.0030 (0.0028; 0.0033)
$\delta_x(3)$	0.0109 (0.0098; 0.0128)	0.00040 (0.00032; 0.00051)	0.0030 (0.0019; 0.0055)	0.0158 (0.0149; 0.0167)
$\beta_2(1)$	0	1474.3 (1155.9; 1833.0)	0	0
$\beta_3(1)$	0	54.1 (33.4; 77.2)	9.3479 (0.9447; 18.633)	0
$\beta_3(2)$	0	0	0.2369 (0.0073; 0.9411)	0
$K_1^P(1,1)$	0.6830 (0.3312; 1.0770)	0.0312 (0.0013; 0.0836)	1.0137 (0.7318; 1.2684)	0.1358 (0.0739; 0.2096)
$K_1^P(1,2)$	0.2414 (−0.2988; 0.7986)	0	−0.0770 (−0.1439; −0.0182)	−0.0461 (−0.1067; −0.0026)
$K_1^P(1,3)$	−0.0237 (−0.1430; 0.0961)	0	0	−0.4445 (−0.6668; −0.2844)
$K_1^P(2,1)$	0.5235 (0.1273; 0.9580)	1.7942 (−4.9003; 8.4018)	−0.1634 (−0.2978; −0.0301)	−0.1661 (−0.1884; −0.1385)
$K_1^P(2,2)$	1.6974 (1.1654; 2.2582)	0.3413 (0.0348; 0.6569)	0.0342 (0.0073; 0.0777)	0.2429 (0.1633; 0.3267)
$K_1^P(2,3)$	0.1225 (0.0022; 0.2450)	1.0003 (−1.6247; 3.6542)	0	−0.0781 (−0.1988; −0.0011)
$K_1^P(3,1)$	0.3282 (−0.0344; 0.6967)	3.2472 (1.9469; 4.7258)	1.4461 (−0.3262; 3.1707)	−0.0033 (−0.0117; −0.0001)
$K_1^P(3,2)$	−0.1186 (−0.6588; 0.4359)	−0.1400 (−0.2254; −0.0681)	−0.0601 (−0.2725; 0.0882)	−0.2937 (−0.3181; −0.2667)
$K_1^P(3,3)$	0.0944 (−0.0121; 0.2100)	1.6085 (1.1046; 2.1460)	0.3833 (0.1446; 0.6255)	1.8887 (1.6339; 2.1414)
$K_0^P(1)$	0.5289 (0.1485; 0.9139)	0.3741 (0.1742; 0.5917)	0.2169 (0.0082; 0.6824)	0.7475 (0.3025; 1.2112)
$K_0^P(2)$	−0.1127 (−0.5192; 0.2938)	0.3850 (−56.1; 55.9)	0.5095 (0.0579; 1.4771)	0.0764 (0.0049; 0.1831)
$K_0^P(3)$	0.2210 (−0.1697; 0.6163)	−0.2580 (−10.90; 10.74)	0.2958 (−2.0638; 2.7795)	0.0427 (0.0011; 0.1596)

Table 5: Model estimates, essentially affine models (part 2). This Table shows parameter estimates along with confidence bands for all 3-factor essentially affine models as defined in section B. The parameters K_0^P and K_1^P are showed for completeness although they are functions of the other parameters and are not estimated. Data used in estimation is Fama and Bliss (1987) monthly zero yields from 1952:6 to 2004:12 and the estimation method is MCMC.

	$A_1(3)$ ext	$A_2(3)$ ext	$A_3(3)$ ext
$K_1^Q(1,1)$	0.0393 (0.0053; 0.0811)	1.1400 (1.0204; 1.2842)	0.2380 (0.1575; 0.3219)
$K_1^Q(1,2)$	0	−0.0457 (−0.0752; −0.0218)	−0.0803 (−0.2062; −0.0027)
$K_1^Q(1,3)$	0	0	−0.0959 (−0.1995; −0.0048)
$K_1^Q(2,1)$	0.0950 (0.0313; 0.1209)	−0.1713 (−0.3081; −0.0112)	−0.1115 (−0.1675; −0.0763)
$K_1^Q(2,2)$	0.4465 (0.3096; 0.5770)	0.0758 (0.0344; 0.1283)	0.2505 (0.1532; 0.4441)
$K_1^Q(2,3)$	1.1754 (1.0330; 1.3771)	0	−0.0437 (−0.1256; −0.0010)
$K_1^Q(3,1)$	0.0235 (0.0027; 0.0478)	3.2571 (3.1596; 3.4077)	−0.01393 (−0.2492; −0.0495)
$K_1^Q(3,2)$	−0.1507 (−0.2083; −0.0944)	−0.3601 (−0.5204; −0.1801)	−1.5435 (−1.8783; −1.2389)
$K_1^Q(3,3)$	0.4294 (0.3313; 0.5361)	0.3221 (0.2517; 0.4396)	1.4639 (1.0538; 1.9630)
$K_0^Q(1)$	1.2579 (0.5305; 2.4484)	0.6003 (0.5022; 0.9402)	0.6566 (0.5043; 1.0310)
$K_0^Q(2)$	0	0.6245 (0.5039; 0.8813)	0.5516 (0.5011; 0.6853)
$K_0^Q(3)$	0	0	1.9349 (0.5268; 4.1132)
$\lambda_2(1,1)$	−0.0326 (−0.1089; 0.0402)	0.1368 (−0.2757; 0.5365)	−0.2922 (−0.4798; −0.1080)
$\lambda_2(1,2)$	0	−0.0041 (−0.0347; 0.0242)	0.0495 (−0.1520; 0.3744)
$\lambda_2(1,3)$	0	0	0.4981 (0.1888; 0.8087)
$\lambda_2(2,1)$	0.0183 (−0.0325; 0.0712)	1.6225 (0.3304; 2.9590)	−0.0679 (−0.1451; 0.0167)
$\lambda_2(2,2)$	0.0290 (−0.1770; 0.2240)	−0.0880 (−0.1876; 0.0105)	−0.4263 (−0.7744; −0.0951)
$\lambda_2(2,3)$	0.6129 (0.1223; 1.1497)	0	0.3264 (0.0933; 0.6674)
$\lambda_2(3,1)$	−0.0819 (−0.1365; −0.0223)	0.4763 (−1.9484; 2.9752)	−0.1179 (−0.2337; −0.0208)
$\lambda_2(3,2)$	−0.0512 (−0.2929; 0.1936)	−0.3019 (−0.5545; −0.0963)	−0.4727 (−0.9290; −0.0444)
$\lambda_2(3,3)$	−1.2157 (−1.6774; −0.7926)	−0.0246 (−0.1969; 0.1468)	0.5495 (0.2107; 0.9574)
σ^2	$5.09e-7$ $(4.70e-7; 5.54e-7)$	$5.25e-7$ $(4.90e-7; 5.63e-7)$	$5.20e-7$ $(4.84e-7; 5.58e-7)$

Table 6: Model estimates, extended affine models (part 1). This Table shows parameter estimates along with confidence bands for all 3-factor essentially affine models as defined in section B. Data used in estimation is Fama and Bliss (1987) monthly zero yields from 1952:6 to 2004:12. The estimation method is MCMC.

	$A_1(3)$ ext	$A_2(3)$ ext	$A_3(3)$ ext
$\lambda_1(1)$	1.0152 $(-0.6526; 2.7914)$	0.0753 $(-0.1420; 0.3477)$	0.4850 $(-0.2748; 1.9402)$
$\lambda_1(2)$	-0.0752 $(-0.7097; 0.5245)$	0.3875 $(-0.1767; 1.4118)$	0.4729 $(-0.0291; 1.5230)$
$\lambda_1(3)$	0.0364 $(-0.5922; 0.6914)$	0.7649 $(-1.8154; 3.4713)$	-0.1342 $(-1.8128; 1.5689)$
δ_0	0.0113 $(0.0078; 0.0149)$	0.0226 $(0.0198; 0.0239)$	0.0017 $(-0.0016; 0.0047)$
$\delta_x(1)$	0.0024 $(0.0023; 0.0026)$	0.0106 $(0.0095; 0.0117)$	-0.0016 $(-0.0021; -0.0012)$
$\delta_x(2)$	0.0074 $(0.0066; 0.0083)$	0.0012 $(0.0007; 0.0015)$	0.0016 $(0.0010; 0.0024)$
$\delta_x(3)$	0.0036 $(0.0022; 0.0053)$	0.0018 $(0.0014; 0.0025)$	0.0043 $(0.0032; 0.0058)$
$\beta_2(1)$	0.0550 $(0.0101; 0.1233)$	0	0
$\beta_3(1)$	0.0460 $(0.0101; 0.1233)$	17.60 $(2.31; 41.50)$	0
$\beta_3(2)$	0	1.846 $(0.481; 3.798)$	0
$K_1^P(1,1)$	0.0719 $(0.0168; 0.1331)$	1.0032 $(0.6080; 1.4148)$	0.5302 $(0.3747; 0.6944)$
$K_1^P(1,2)$	0	-0.0415 $(-0.0762; -0.0140)$	-0.1298 $(-0.4398; -0.0039)$
$K_1^P(1,3)$	0	0	-0.5939 $(-0.9026; -0.3109)$
$K_1^P(2,1)$	0.0767 $(0.0201; 0.1313)$	-1.7937 $(-3.1157; -0.5323)$	-0.0436 $(-0.1214; -0.0018)$
$K_1^P(2,2)$	0.4175 $(0.2358; 0.6045)$	0.1638 $(0.0749; 0.2606)$	0.6768 $(0.3098; 1.0599)$
$K_1^P(2,3)$	0.5625 $(0.0538; 1.0734)$	0	-0.3701 $(-0.7065; -0.1287)$
$K_1^P(3,1)$	0.1054 $(0.0326; 0.1725)$	2.7808 $(0.2578; 5.2090)$	-0.0215 $(-0.0746; -0.0005)$
$K_1^P(3,2)$	-0.0994 $(-0.3563; 0.1339)$	-0.0583 $(-0.2439; 0.1089)$	-1.0708 $(-1.5926; -0.6526)$
$K_1^P(3,3)$	1.6451 $(1.2099; 2.1238)$	0.3467 $(0.1764; 0.5238)$	0.9144 $(0.5486; 1.3692)$
$K_0^P(1)$	2.2731 $(1.0089; 3.8133)$	0.6757 $(0.5083; 1.0301)$	1.1416 $(0.5175; 2.5748)$
$K_0^P(2)$	-0.0752 $(-0.7097; 0.5245)$	1.0119 $(0.5188; 2.0401)$	1.0245 $(0.5165; 2.0921)$
$K_0^P(3)$	0.0364 $(-0.5922; 0.6914)$	0.7649 $(-1.8154; 3.4713)$	1.8007 $(0.5679; 3.7045)$

Table 7: Model estimates, extended affine models (part 2). This Table shows parameter estimates along with confidence bands for all 3-factor essentially affine models as defined in section B. The parameters K_0^P and K_1^P are showed for completeness although they are functions of the other parameters and are not estimated. Data used in estimation is Fama and Bliss (1987) monthly zero yields from 1952:6 to 2004:12 and the estimation method is MCMC.

n	2	3	4	5
Actual	−0.775	−1.131	−1.520	−1.494
$A_0(3)ess$				
Implied	−0.311 (−1.155; 0.464)	−0.415 (−1.382; 0.473)	−0.531 (−1.588; 0.447)	−0.657 (−1.811; 0.406)
Simulated	−0.206 (−1.089; 0.497)	−0.316 (−1.338; 0.519)	−0.447 (−1.561; 0.479)	−0.591 (−1.798; 0.417)
$A_1(3)ess$				
Implied	−0.119 (−0.950; 0.706)	−0.166 (−0.877; 0.702)	−0.066 (−0.767; 0.777)	−0.100 (−0.647; 0.879)
Simulated	0.302 (−0.364; 0.887)	0.243 (−0.425; 0.892)	0.262 (−0.430; 0.971)	0.314 (−0.418; 1.075)
$A_2(3)ess$				
Implied	−0.007 (−0.927; 0.910)	0.029 (−0.818; 0.977)	0.100 (−0.704; 1.062)	0.185 (−0.598; 1.137)
Simulated	0.558 (−0.087; 1.169)	0.587 (−0.124; 1.231)	0.636 (−0.110; 1.338)	0.690 (−0.084; 1.415)
$A_3(3)ess$				
Implied	0.567 (−0.604; 1.026)	0.631 (−0.482; 1.078)	0.728 (−0.312; 1.135)	0.835 (−0.135; 1.194)
Simulated	1.080 (0.796; 1.323)	1.158 (0.827; 1.424)	1.236 (0.870; 1.530)	1.309 (0.946; 1.602)
$A_1(3)ext$				
Implied	0.070 (−0.479; 0.506)	−0.057 (−0.575; 0.383)	−0.067 (−0.571; 0.377)	−0.005 (−0.504; 0.441)
Simulated	0.231 (−0.293; 0.649)	0.089 (−0.443; 0.541)	0.071 (−0.502; 0.532)	0.134 (−0.441; 0.598)
$A_2(3)ext$				
Implied	0.120 (−0.484; 0.682)	0.086 (−0.509; 0.677)	0.095 (−0.498; 0.713)	0.130 (−0.468; 0.768)
Simulated	0.381 (−0.138; 0.872)	0.340 (−0.232; 0.878)	0.337 (−0.289; 0.924)	0.359 (−0.302; 0.981)
$A_3(3)ext$				
Implied	0.191 (−0.270; 0.670)	0.135 (−0.370; 0.679)	0.084 (−0.454; 0.676)	0.030 (−0.538; 0.669)
Simulated	0.282 (−0.184; 0.765)	0.233 (−0.326; 0.826)	0.192 (−0.421; 0.851)	0.147 (−0.502; 0.846)

Table 8: Model-implied Campbell-Shiller regression coefficients. This table shows the regression coefficients from the regressions $Y(t+1,n-1) - Y(t,n) = const + \phi_n[\frac{Y(t,n)-Y(t,1)}{n-1}] + residual$ where n and t are measured in years. The first set is the model-implied coefficients calculated according to Appendix D. The second set takes into account finite-sample bias by simulating as explained in the text.

n	2	3	4	5
Actual	1	1	1	1
$A_0(3)$ ess	0.525 $(-0.237; 1.362)$	1.099 $(-0.171; 2.616)$	0.881 $(-0.452; 2.542)$	1.068 $(-0.337; 2.887)$
$A_1(3)$ ess	-0.150 $(-0.677; 0.418)$	0.060 $(-0.757; 1.032)$	-0.299 $(-1.107; 0.762)$	-0.292 $(-1.106; 0.880)$
$A_2(3)$ ess	-0.328 $(-0.916; 0.342)$	-0.333 $(-1.235; 0.802)$	-0.651 $(-1.558; 0.590)$	-0.588 $(-1.536; 0.820)$
$A_3(3)$ ess	-0.672 $(-0.748; -0.590)$	-0.921 $(-1.095; -0.744)$	-1.282 $(-1.536; -1.032)$	-1.276 $(-1.582; -0.975)$
$A_1(3)$ ext	0.058 $(-0.289; 0.416)$	0.490 $(-0.046; 1.071)$	0.124 $(-0.398; 0.728)$	0.088 $(-0.425; 0.714)$
$A_2(3)$ ext	-0.086 $(-0.554; 0.441)$	0.137 $(-0.615; 1.001)$	-0.125 $(-0.896; 0.774)$	-0.022 $(-0.828; 0.937)$
$A_3(3)$ ext	0.033 $(-0.383; 0.419)$	0.310 $(-0.343; 0.900)$	0.037 $(-0.587; 0.588)$	0.184 $(-0.441; 0.721)$

Table 9: Risk-premium adjusted Campbell-Shiller regression coefficients. This table shows the slope coefficents from the regressions $Y(t+1,n-1) - Y(t,n) + D^*(t+1,n) = const + \phi_n[\frac{Y(t,n)-Y(t,1)}{n-1}] + residual$ where n and t are measured in years. In the regression actual yields and model-implied risk premia D^* defined in equation (9) are used.

n	1	2	3	4	5
Actual	5.60	5.81	5.98	6.11	6.19
$A_0(3)$ ess					
mean	6.12	6.37	6.57	6.71	6.81
median	5.86	6.09	6.31	6.45	6.55
	(1.22; 13.75)	(1.48; 14.01)	(1.63; 14.17)	(1.74; 14.25)	(1.83; 14.28)
$A_1(3)$ ess					
mean	9.56	10.06	10.44	10.71	10.88
median	4.63	4.80	4.92	5.02	5.10
	(2.29; 39.29)	(2.48; 38.98)	(2.72; 39.46)	(2.82; 40.39)	(2.84; 40.90)
$A_2(3)$ ess					
mean	28.69	30.74	32.17	33.10	33.63
median	5.25	5.49	5.66	5.76	5.83
	(2.44; 49.77)	(2.57; 52.85)	(2.70; 53.60)	(2.76; 54.09)	(2.76; 55.24)
$A_3(3)$ ess					
mean	15.42	15.89	16.23	16.43	16.52
median	5.77	6.02	6.22	6.39	6.53
	(2.59; 49.81)	(2.71; 50.98)	(2.80; 52.04)	(2.88; 52.74)	(2.94; 53.08)
$A_1(3)$ ext					
mean	6.83	7.09	7.29	7.44	7.53
median	6.27	6.54	6.72	6.89	7.02
	(4.15; 13.18)	(4.37; 13.40)	(4.54; 13.56)	(4.67; 13.66)	(4.78; 13.71)
$A_2(3)$ ext					
mean	6.39	6.68	6.90	7.06	7.18
median	5.97	6.23	6.46	6.54	6.64
	(4.50; 11.19)	(4.68; 11.89)	(4.84; 12.34)	(4.95; 12.51)	(5.04; 12.67)
$A_3(3)$ ext					
mean	6.23	6.46	6.65	6.79	6.88
median	6.00	6.24	6.42	6.55	6.64
	(4.04; 9.61)	(4.29; 9.86)	(4.50; 10.10)	(4.67; 10.26)	(4.80; 10.36)

Table 10: Unconditional mean of yields. The first line in this Table shows the unconditional mean in percent of the yields in the data where n denotes maturity. The next lines show the model-implied unconditional mean, median, and confidence bands of yields for the estimated models. These are calculated on basis of simulated yields as explained in the text.

n	1	2	3	4	5
Actual	49.3	43.2	40.1	38.8	36.2
$A_0(3)$ ess					
mean	48.8	43.6	39.8	37.2	35.4
median	48.7	43.6	39.8	37.2	35.4
	(46.3; 51.6)	(41.4; 45.9)	(37.9; 41.9)	(35.3; 39.2)	(33.6; 37.3)
$A_1(3)$ ess					
mean	47.2	42.5	39.9	37.9	36.1
median	35.7	32.3	30.4	28.8	27.5
	(18.1; 134.2)	(16.4; 121.1)	(15.4; 113.6)	(14.7; 107.5)	(14.0; 102.2)
$A_2(3)$ ess					
mean	54.6	48.9	45.6	42.8	40.4
median	39.0	35.1	32.6	30.7	28.9
	(15.3; 159.8)	(13.1; 141.3)	(12.0; 130.5)	(11.1; 122.9)	(10.4; 116.1)
$A_3(3)$ ess					
mean	63.6	56.8	52.9	49.9	47.5
median	48.6	43.5	40.5	38.2	36.3
	(27.8; 158.8)	(24.9; 142.3)	(23.1; 132.6)	(21.8; 125.0)	(20.6; 118.7)
$A_1(3)$ ext					
mean	45.6	40.5	37.6	35.3	33.4
median	44.2	39.4	36.6	34.5	32.4
	(35.3; 67.1)	(31.5; 57.7)	(29.5; 52.7)	(27.6; 49.6)	(26.0; 47.0)
$A_2(3)$ ext					
mean	48.3	42.9	40.0	37.9	36.2
median	46.6	41.4	38.6	36.6	35.0
	(36.9; 71.1)	(32.1; 63.9)	(29.9; 60.0)	(28.3; 57.0)	(27.0; 54.5)
$A_3(3)$ ext					
mean	45.9	42.8	40.3	37.9	35.6
median	45.4	42.1	39.6	37.2	35.0
	(35.8; 58.8)	(33.0; 54.9)	(31.2; 51.5)	(29.5; 48.2)	(27.9; 45.3)

Table 11: Unconditional volatility of yields. The first lines in this Table shows the unconditional volatility (standard deviation) in basispoints of monthly yield changes in the data where n denotes maturity. The next lines show the model-implied unconditional mean, median, and confidence bands of volatility of monthly yield changes for the estimated models. These are calculated on basis of simulated yields as explained in the text.

n	1	2	3	4	5
Actual					
level	0.110	0.071	0.057	0.052	0.044
slope	−0.142	−0.079	−0.032	−0.020	0.016
curvature	0.271	0.108	0.166	0.078	0.126
$A_0(3)$					
level	0.000	0.000	0.000	0.000	0.000
	(−0.002; 0.002)	(−0.001; 0.001)	(−0.001; 0.001)	(−0.001; 0.001)	(−0.001; 0.001)
slope	0.000	0.000	0.000	0.000	0.000
	(−0.009; 0.010)	(−0.007; 0.008)	(−0.006; 0.006)	(−0.005; 0.006)	(−0.005; 0.005)
curvature	0.001	0.001	0.001	0.000	0.000
	(−0.021; 0.025)	(−0.018; 0.020)	(−0.016; 0.016)	(−0.013; 0.014)	(−0.011; 0.012)
$A_1(3)$ ess					
level	0.045	0.036	0.032	0.029	0.026
	(0.040; 0.051)	(0.032; 0.041)	(0.029; 0.036)	(0.026; 0.032)	(0.024; 0.029)
slope	0.065	0.052	0.045	0.041	0.037
	(0.049; 0.087)	(0.039; 0.070)	(0.034; 0.062)	(0.030; 0.055)	(0.028; 0.049)
curvature	0.117	0.093	0.081	0.072	0.065
	(0.082; 0.159)	(0.065; 0.128)	(0.054; 0.112)	(0.047; 0.101)	(0.042; 0.091)
$A_2(3)$ ess					
level	0.043	0.035	0.030	0.026	0.023
	(0.033; 0.052)	(0.027; 0.041)	(0.023; 0.035)	(0.020; 0.031)	(0.018; 0.028)
slope	0.128	0.096	0.080	0.068	0.059
	(0.070; 0.169)	(0.055; 0.127)	(0.050; 0.107)	(0.044; 0.092)	(0.039; 0.080)
curvature	0.359	0.255	0.204	0.169	0.142
	(0.141; 0.513)	(0.106; 0.365)	(0.097; 0.293)	(0.085; 0.245)	(0.075; 0.207)
$A_3(3)$ ess					
level	0.051	0.039	0.033	0.029	0.027
	(0.048; 0.054)	(0.037; 0.041)	(0.031; 0.035)	(0.028; 0.031)	(0.025; 0.028)
slope	0.020	−0.006	−0.012	−0.012	−0.011
	(0.007; 0.038)	(−0.016; 0.007)	(−0.021; −0.002)	(−0.020; −0.005)	(−0.018; −0.005)
curvature	0.130	0.013	−0.016	−0.023	−0.023
	(0.096; 0.171)	(−0.014; 0.042)	(−0.040; 0.007)	(−0.044; −0.004)	(−0.042; −0.006)
$A_1(3)$ ext					
level	0.032	0.025	0.022	0.019	0.018
	(0.024; 0.040)	(0.019; 0.032)	(0.016; 0.027)	(0.015; 0.024)	(0.014; 0.022)
slope	0.038	0.030	0.026	0.022	0.021
	(0.024; 0.055)	(0.019; 0.042)	(0.016; 0.036)	(0.014; 0.031)	(0.013; 0.028)
curvature	0.079	0.061	0.052	0.046	0.042
	(0.045; 0.118)	(0.037; 0.091)	(0.031; 0.077)	(0.028; 0.067)	(0.025; 0.059)
$A_2(3)$ ext					
level	0.047	0.037	0.032	0.029	0.026
	(0.039; 0.055)	(0.031; 0.043)	(0.027; 0.037)	(0.025; 0.033)	(0.023; 0.030)
slope	0.136	0.097	0.080	0.070	0.062
	(0.095; 0.181)	(0.066; 0.130)	(0.056; 0.108)	(0.049; 0.093)	(0.044; 0.083)
curvature	0.370	0.238	0.186	0.156	0.135
	(0.201; 0.567)	(0.122; 0.384)	(0.100; 0.294)	(0.088; 0.242)	(0.079; 0.205)
$A_3(3)$ ext					
level	0.039	0.033	0.029	0.025	0.022
	(0.030; 0.050)	(0.025; 0.043)	(0.022; 0.038)	(0.019; 0.034)	(0.017; 0.030)
slope	0.024	0.004	−0.003	−0.005	−0.005
	(0.006; 0.049)	(−0.011; 0.026)	(−0.016; 0.014)	(−0.016; 0.009)	(−0.014; 0.006)
curvature	0.062	0.002	−0.017	−0.022	−0.022
	(0.025; 0.107)	(−0.029; 0.038)	(−0.044; 0.012)	(−0.045; 0.001)	(−0.041; −0.004)

Table 12: *Volatility regression.* This table shows for the affine models the coefficents from the regressions $[Y(t+1,n)-Y(t,n)]^2 = const + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5)-Y(t,1)] + \phi_n(3)[Y(t,5)+Y(t,1)-2Y(t,3)] + residual$ where t is measured in months, n in years, and Y in percent. The regression coefficients are calculated using simulated yields (in percent) .

Whole sample period 1952:6 to 2004:12

n	1	2	3	4	5
level	0.1095***	0.0713***	0.0565***	0.0519***	0.0438***
	(0.0202)	(0.0132)	(0.0102)	(0.0075)	(0.0058)
slope	−0.1415	−0.0785	−0.0316	−0.0196	0.0156
	(0.0960)	(0.0631)	(0.0479)	(0.0362)	(0.0286)
curvature	0.2712	0.1082	0.1657	0.0776	0.1262*
	(0.1968)	(0.1304)	(0.0974)	(0.0755)	(0.0603)

Period before the Fed experiment 1952:6 to 1979:9

n	1	2	3	4	5
level	0.0465***	0.0285***	0.0266***	0.0253***	0.0183**
	(0.0097)	(0.0068)	(0.0067)	(0.0061)	(0.0056)
slope	0.0551	0.0368	0.0272	0.0236	0.0201
	(0.0483)	(0.0349)	(0.0326)	(0.0303)	(0.0263)
curvature	0.1330	0.0014	0.0107	0.0240	0.0539
	(0.0795)	(0.0583)	(0.0531)	(0.0498)	(0.0419)

Period after the Fed experiment 1982:11 to 2004:12

n	1	2	3	4	5
level	0.0306***	0.0195***	0.0174***	0.0198***	0.0180***
	(0.0051)	(0.0048)	(0.0046)	(0.0049)	(0.0049)
slope	0.0254	0.0316	0.0511**	0.0639**	0.0605**
	(0.0209)	(0.0195)	(0.0191)	(0.0199)	(0.0200)
curvature	0.0368	0.0577	0.0750	0.1268*	0.1116
	(0.0601)	(0.0564)	(0.0560)	(0.0584)	(0.0575)

Table 13: *Volatility regression in subperiods.* This table shows the coefficients from the same regression as Table 3 except that the regression is split into the period before the Fed Experiment and after the Fed Experiment. The regression is $[Y(t+1,n) - Y(t,n)]^2 = const + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5) - Y(t,1)] + \phi_n(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] + residual$ and in parenthesis are shown Hansen and Hodrick (1980) standard errors with 6 lags and significance at the 5%, 1%, or 0.1% level is denoted by $*, **$, or $***$.

Whole sample period 1952:6 to 2004:12

	$A_0(3)ess$	$A_1(3)ess$	$A_2(3)ess$	$A_3(3)ess$	$A_1(3)ext$	$A_2(3)ext$	$A_3(3)ext$
1-year	0.0	70.2	69.3	73.2	72.7	69.9	73.0
2-year	0.0	74.3	73.0	72.3	75.3	74.0	74.0
3-year	0.0	78.4	78.0	74.3	78.9	79.0	76.2
4-year	0.0	81.0	80.4	76.1	81.1	81.4	78.0
5-year	0.0	79.0	79.5	73.2	78.8	80.2	74.9
average	0.0	76.6	76.1	73.8	77.4	76.9	75.2

Period before the Fed experiment 1952:6 to 1979:9

	$A_0(3)ess$	$A_1(3)ess$	$A_2(3)ess$	$A_3(3)ess$	$A_1(3)ext$	$A_2(3)ext$	$A_3(3)ext$
1-year	0.0	77.6	75.3	72.0	77.0	76.2	75.6
2-year	0.0	72.7	68.3	67.2	70.7	70.9	70.4
3-year	0.0	75.9	73.6	71.5	74.3	76.3	73.8
4-year	0.0	77.2	75.9	72.8	75.7	78.5	74.9
5-year	0.0	67.7	68.6	62.5	66.1	70.3	64.2
average	0.0	74.2	72.3	69.2	72.8	74.4	71.8

Period after the Fed experiment 1982:11 to 2004:12

	$A_0(3)ess$	$A_1(3)ess$	$A_2(3)ess$	$A_3(3)ess$	$A_1(3)ext$	$A_2(3)ext$	$A_3(3)ext$
1-year	0.0	74.8	56.0	74.1	74.9	61.9	76.1
2-year	0.0	66.5	54.3	67.5	66.6	60.6	68.4
3-year	0.0	72.5	61.0	76.0	73.1	67.7	76.0
4-year	0.0	61.9	52.7	66.4	62.8	58.2	66.4
5-year	0.0	54.1	46.0	59.5	55.1	50.6	59.3
average	0.0	66.0	54.0	68.7	66.5	59.8	69.2

Table 14: *Correlation between conditional volatility and an EGARCH estimate.* This table shows the correlation between model-implied monthly conditional volatility and an EGARCH(1,1) estimate of montly conditional volatility.

Period with daily data available 1961:7 to 2004:12

n	1	2	3	4	5
Actual	34.5	33.2	32.8	31.9	31.1
$A_1(3)ess$	45.1	45.1	45.1	45.1	45.1
	(9.5; 75.8)	(9.4; 76.2)	(9.3; 76.5)	(9.1; 76.4)	(8.9; 76.2)
$A_2(3)ess$	19.8	21.0	22.2	23.4	24.8
	(0.8; 50.4)	(0.8; 52.4)	(1.2; 54.6)	(1.2; 56.7)	(1.2; 58.1)
$A_3(3)ess$	62.7	62.8	63.2	63.3	63.2
	(38.8; 82.8)	(38.1; 82.8)	(39.1; 83.1)	(39.0; 83.1)	(39.0; 83.1)
$A_1(3)ext$	28.7	28.6	28.5	28.6	29.1
	(7.0; 56.6)	(5.8; 56.7)	(5.8; 56.5)	(6.4; 56.7)	(6.3; 57.6)
$A_2(3)ext$	21.9	25.3	27.6	29.2	30.4
	(1.9; 51.8)	(2.6; 55.2)	(3.5; 58.0)	(4.0; 58.8)	(4.8; 59.5)
$A_3(3)ext$	43.7	46.2	47.7	48.6	49.1
	(21.2; 65.8)	(25.5; 67.9)	(26.4; 69.2)	(27.8; 70.0)	(28.7; 70.3)

Period after Fed experiment 1982:11 to 2004:12

n	1	2	3	4	5
Actual	5.2	5.5	6.9	8.1	8.8
$A_1(3)ess$	36.1	36.1	36.1	36.1	36.1
	(3.0; 75.4)	(2.7; 75.4)	(2.7; 75.4)	(2.9; 74.9)	(3.1; 74.6)
$A_2(3)ess$	19.2	18.7	18.7	19.1	19.6
	(0.4; 50.8)	(0.6; 52.6)	(0.8; 53.7)	(0.9; 54.9)	(0.7; 55.8)
$A_3(3)ess$	55.2	55.2	55.7	55.9	55.7
	(27.9; 81.1)	(28.1; 81.6)	(28.9; 81.9)	(29.5; 82.0)	(28.8; 81.8)
$A_1(3)ext$	21.2	21.2	21.0	21.2	21.7
	(2.2; 47.5)	(2.4; 48.6)	(2.6; 48.6)	(2.7; 50.0)	(2.8; 50.6)
$A_2(3)ext$	19.5	21.3	22.9	24.0	25.0
	(0.8; 49.3)	(0.1; 51.0)	(1.8; 54.0)	(2.1; 55.9)	(2.7; 56.7)
$A_3(3)ext$	35.4	38.2	40.0	41.0	41.6
	(12.2; 61.6)	(15.7; 63.3)	(17.8; 65.4)	(18.7; 66.0)	(19.4; 66.4)

Table 15: Testing the fundamental affine yield variation spanning condition. This table shows the R^2's (in percent) from regressing monthly realized volatility (calculated using daily data) on average monthly level, slope, and curvature. 'Actual' are the R^2's from historical data. Model-implied mean R^2's and 95% confidence bands are calculated by simulating 500 data sets of equal length as the historical data. In the simulations mean parameter estimates are used.

n	2	3	4	5
level	0.00183	0.00138	0.00121	0.00127
	(0.00031)	(0.00023)	(0.00019)	(0.00020)
slope	0.00070	0.00094	0.00099	0.00132
	(0.00183)	(0.00141)	(0.00116)	(0.00112)
curvature	0.00529	0.00291	0.00206	0.00248
	(0.00456)	(0.00361)	(0.00301)	(0.00280)
ARCH	0.76347	0.78226	0.77349	0.74372
	(0.02870)	(0.02855)	(0.02885)	(0.02913)

Table 16: Volatility regression. This table shows the slope coefficents from the regressions $[Y(t+1,n-1)-E_t(Y(t+1,n-1))]^2 = const+\phi_n(1)[Y(t,5)]+\phi_n(2)[Y(t,5)-Y(t,1)]+\phi_n(3)[Y(t,5)+Y(t,1)-2Y(t,3)]+\phi_n(4)[Y(t,n-1)-E_{t-1}(Y(t,n-1))]^2+residual$ where n and t are measured in years. In parentheses are shown Hansen and Hodrick (1980) standard errors with 12 lags. Data: Fama and Bliss (1987) monthly observations from 1952:6 to 2004:12.

n	2	3	4	5
Actual				
level	0.183	0.138	0.121	0.127
slope	0.070	0.094	0.099	0.132
curvature	0.529	0.291	0.206	0.248
ARCH	0.763	0.782	0.773	0.733
$A_0(3)$				
level	0.000	0.000	0.000	0.000
	$(-0.010; 0.011)$	$(-0.008; 0.008)$	$(-0.007; 0.006)$	$(-0.006; 0.005)$
slope	-0.001	-0.001	0.000	0.000
	$(-0.059; 0.054)$	$(-0.047; 0.043)$	$(-0.038; 0.036)$	$(-0.033; 0.032)$
curvature	-0.003	-0.002	-0.001	-0.001
	$(-0.139; 0.135)$	$(-0.114; 0.112)$	$(-0.095; 0.096)$	$(-0.081; 0.085)$
ARCH	0.809	0.809	0.808	0.808
	$(0.790; 0.828)$	$(0.790; 0.828)$	$(0.791; 0.827)$	$(0.791; 0.825)$
$A_1(3)$ ess				
level	0.088	0.074	0.066	0.059
	$(0.069; 0.113)$	$(0.060; 0.092)$	$(0.053; 0.080)$	$(0.048; 0.071)$
slope	0.059	0.049	0.047	0.046
	$(-0.013; 0.121)$	$(-0.008; 0.096)$	$(-0.004; 0.085)$	$(0.003; 0.080)$
curvature	0.125	0.083	0.070	0.068
	$(-0.016; 0.272)$	$(-0.035; 0.203)$	$(-0.030; 0.172)$	$(-0.021; 0.150)$
ARCH	0.787	0.786	0.785	0.785
	$(0.766; 0.811)$	$(0.766; 0.808)$	$(0.762; 0.805)$	$(0.763; 0.806)$
$A_2(3)$ ess				
level	0.076	0.064	0.056	0.050
	$(0.053; 0.101)$	$(0.045; 0.084)$	$(0.039; 0.072)$	$(0.035; 0.063)$
slope	0.130	0.105	0.091	0.081
	$(0.051; 0.221)$	$(0.041; 0.174)$	$(0.037; 0.152)$	$(0.034; 0.138)$
curvature	0.285	0.217	0.180	0.153
	$(0.096; 0.531)$	$(0.060; 0.393)$	$(0.046; 0.325)$	$(0.040; 0.281)$
ARCH	0.794	0.797	0.798	0.798
	$(0.765; 0.820)$	$(0.771; 0.823)$	$(0.771; 0.821)$	$(0.771; 0.821)$
$A_3(3)$ ess				
level	0.080	0.067	0.059	0.053
	$(0.065; 0.095)$	$(0.056; 0.079)$	$(0.050; 0.069)$	$(0.045; 0.062)$
slope	-0.041	-0.035	-0.026	-0.017
	$(-0.119; 0.026)$	$(-0.099; 0.022)$	$(-0.084; 0.025)$	$(-0.068; 0.028)$
curvature	-0.159	-0.149	-0.127	-0.103
	$(-0.345; 0.015)$	$(-0.322; 0.001)$	$(-0.277; 0.004)$	$(-0.245; 0.015)$
ARCH	0.791	0.796	0.798	0.799
	$(0.773; 0.807)$	$(0.777; 0.813)$	$(0.779; 0.813)$	$(0.780; 0.814)$

Table 17: Fitted volatility regressions. This table shows for the essentially affine models the regression coefficients from the regressions $[Y(t+1,n-1)-E_t(Y(t+1,n-1))]^2 = const+\phi_n(1)[Y(t,5)]+\phi_n(2)[Y(t,5)-Y(t,1)]+\phi_n(3)[Y(t,5)+Y(t,1)-2Y(t,3)]+\phi_n(4)[Y(t,n-1)-E_{t-1}(Y(t,n-1))]^2+residual$ where n and t are measured in years. The regression coefficients are calculated using simulated yields (in percent) .

n	2	3	4	5
Actual				
level	0.183	0.138	0.121	0.127
slope	0.070	0.094	0.099	0.132
curvature	0.529	0.291	0.206	0.248
ARCH	0.763	0.782	0.773	0.733
$A_1(3)$ ext				
level	0.057	0.048	0.042	0.037
	$(0.044; 0.075)$	$(0.037; 0.063)$	$(0.032; 0.055)$	$(0.029; 0.049)$
slope	0.041	0.034	0.031	0.029
	$(-0.004; 0.085)$	$(-0.004; 0.071)$	$(-0.001; 0.062)$	$(0.000; 0.057)$
curvature	0.104	0.079	0.068	0.062
	$(-0.003; 0.205)$	$(-0.015; 0.163)$	$(-0.019; 0.138)$	$(-0.013; 0.123)$
ARCH	0.799	0.797	0.795	0.796
	$(0.784; 0.812)$	$(0.781; 0.812)$	$(0.778; 0.812)$	$(0.778; 0.813)$
$A_2(3)$ ext				
level	0.077	0.066	0.059	0.053
	$(0.061; 0.095)$	$(0.053; 0.081)$	$(0.047; 0.071)$	$(0.043; 0.064)$
slope	0.134	0.115	0.106	0.099
	$(0.070; 0.195)$	$(0.060; 0.172)$	$(0.059; 0.157)$	$(0.056; 0.145)$
curvature	0.317	0.253	0.223	0.204
	$(0.146; 0.503)$	$(0.105; 0.402)$	$(0.091; 0.360)$	$(0.083; 0.329)$
ARCH	0.795	0.797	0.797	0.797
	$(0.780; 0.810)$	$(0.783; 0.810)$	$(0.785; 0.810)$	$(0.785; 0.810)$
$A_3(3)$ ext				
level	0.063	0.055	0.048	0.043
	$(0.045; 0.085)$	$(0.040; 0.074)$	$(0.035; 0.065)$	$(0.031; 0.058)$
slope	-0.036	-0.039	-0.034	-0.028
	$(-0.089; 0.014)$	$(-0.083; 0.006)$	$(-0.075; 0.007)$	$(-0.066; 0.010)$
curvature	-0.086	-0.105	-0.099	-0.086
	$(-0.219; 0.021)$	$(-0.211; -0.015)$	$(-0.186; -0.024)$	$(-0.159; -0.017)$
ARCH	0.791	0.788	0.787	0.786
	$(0.773; 0.807)$	$(0.770; 0.808)$	$(0.768; 0.808)$	$(0.767; 0.808)$

Table 18: Fitted volatility regressions. This table shows for the extended affine models the slope coefficients from the regressions $[Y(t{+}1, n{-}1) - E_t(Y(t{+}1, n{-}1))]^2 = const + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5) - Y(t,1)] + \phi_n(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] + \phi_n(4)[Y(t, n{-}1) - E_{t-1}(Y(t, n{-}1))]^2 + residual$ where n and t are measured in years. The regression coefficients are calculated using simulated yields (in percent) .