Quarterly Journal of Finance
Vol. 6, No. 2 (2016) 1650009 (56 pages)

DOI: 10.1142/S2010139216500099

Can Affine Models Match the Moments in Bond Yields?*

Peter Feldhütter
London Business School
Regent's Park, London
NW1 4SA, United Kingdom
pfeldhutter@london.edu

Accepted 9 September 2015
Published 29 October 2015

This paper examines the ability of three-factor affine term structure models with essentially, extended, and semi-affine risk premium specifications to capture the dynamics of bond excess returns, yield volatility and higher order moments in yields. Extended affine models can best capture the time-variation in excess returns and yield volatility simultaneous. However, none of the three-factor models can fully match bond return predictability and yield volatility jointly. Extended affine models are more restricted in the ability to price bonds because of necessary parameter restrictions — the so-called Feller condition — and essentially affine and semi-affine models are therefore better suited for pricing purposes.

Keywords: Affine term structure models; market price of risk; time-varying risk premium; time-varying volatility; Feller condition.

1. Introduction

Empirical evidence suggests that risk premia and volatility in US interest rates are time-varying. The excess return on a bond is documented by among others Campbell and Shiller (1991) 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. Duffee (2002) and Dai and Singleton (2002) find that standard affine term structure models

*Previously circulated under the title "Can Affine Models Capture the Dynamics of Risk Premia and Volatility in Bond Yields?".

with Duffee (2002)'s essentially affine risk premium specification can fit one of these empirical facts but have difficulty in fitting both.

I examine whether the tension between fitting the first and second conditional moments is less pronounced in extended affine models proposed by Cheridito *et al.* (2007) and semi-affine models proposed by Duarte (2004). I investigate models with exactly three latent factors. The reason for choosing three factors is that Litterman and Scheinkman (1991) show that three factors explain almost all the variation in the yield curve while Duffee (2010) shows that Sharpe ratios in affine models with more than three factors are "astronomically high".

Using Campbell and Shiller (1991) regression coefficients as a benchmark, I show that extended models capture risk premia better than essentially and semi-affine models. Although none of the models with stochastic volatility match a purely Gaussian model, extended models with stochastic volatility outperform their essentially and semi-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 extended models the ability is similar across models with one, two, and three volatility factors and all three extended models do as well or better than the essentially affine model with one volatility factor ($A_1(3)$ model). Although semi-affine models nest essentially affine models, the added flexibility does not help in capturing risk premia.

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. Comparing model-implied volatility with historical volatility I show that none of the models capture the high volatility during the Fed experiment in 1979–1982, but apart from this period the broad trends in volatility are matched. Also, 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 leading to the conclusion that the $A_1(3)$ models capture the dynamics of volatility slightly better in these subperiods. I also document a tension in essentially and extended affine models between fitting yields in both the time series and cross section. The distribution of yields in extended models is more in accordance with the historical distribution than in essentially affine models: skewness and kurtosis of yields in extended models are 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 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 distribution from the risk-neutral distribution, while in the extended 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 extended models are less successful in fitting the cross section of yields. The reason is that extended models are only well defined if additional restrictions are imposed in the model, namely the so-called Feller restrictions which will be described in detail later. These restrictions limit the possible yield curve shapes that extended models can generate and the restrictions are binding. Therefore, the cross-sectional fit in extended models is worse than in essentially affine models.

The semi-affine models of Duarte (2004) solve the tension in fitting both the time series and cross-sectional properties of yields. The semi-affine models (1) have the ability to fit the higher order moments of yields under the historical measure, (2) have the ability to generate a wide variety of yield curve shapes under the risk-neutral measure, and (3) have the same degree of predictability as essentially affine models. Thus, the added parameters in semi-affine models improve their fit relative to essentially affine models, but at the cost of non-affine dynamics of the factors under the actual measure.

Finally, the results show that the unconditional mean of yields is hard to pin down with reasonable precision. For example, the average five-year yield is 6.19% while the 95% confidence band in the essentially affine $A_1(3)$ model is $[2.84\%, 40.90\%]$. Thus, the average yield curve is not a moment that is useful to statistically discriminate between models.

The paper is organized as follows. Section 2 describes the features in the US term structure that affine models should match. The affine framework is setup in Sec. 3 and the estimation methodology Markov Chain Monte Carlo (MCMC) is explained in Sec. 4. The fit of essentially, extended, and semi-affine three-factor affine models is examined in Sec. 5 and Sec. 6 concludes.

2. Features of the US Term Structure

I compare three-factor affine term structure models across a set of US term structure "features" that are easily interpretable and has proven difficult for term structure models to match. Earlier studies have used these features to compare completely affine, essentially affine, and/or quadratic term structure models.

The data are month-end (continuously compounded) 1, 2, 3, 4, and 5 years zero-coupon yields for the period 1952:6 to 2004:12 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 (CRSP) and are used both in this section to illustrate the US term structure features and in the later estimation of affine models. Below, I describe the moments in the data across which I compare affine models.

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

Expected excess returns in US Treasury bonds vary across time and maturity. Fama (1984) and Campbell and Shiller (1991) show that expected excess returns are positively related to the slope of the yield curve and the effect increases with the maturity of the bond. Since their seminal papers a number of other factors predicting excess returns have been documented (see for example Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), Cooper and Priestley (2009), Joslin *et al.* (2014), and Cieslak and Povala (2015)), but I use the slope of the yield curve as a test because Dai and Singleton (2002) show that it is a hard test for affine models and because this predictor has passed the test of time and shown to be a consistent predictor of excess returns.[1]

Campbell and Shiller (1991) run the regression

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

where $Y(t, n)$ is the n-year zero-coupon yield at time t and the coefficients are given in Panel B in Table 1. With constant risk premia the expectation theory predicts that ϕ_n is equal to one for all maturities, but the actual coefficients are negative and increasingly so with maturity.

[1] See Duffee (2012) for a discussion of the out-of-sample properties of several of the recently documented factors.

Table 1. Yield Curve Statistics. Panel A shows the unconditional mean, skewness, and kurtosis 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. Panel B shows the slope coefficients from the regressions $Y(t+1,n-1) - Y(t,n) = \text{const} + \phi_n[\frac{Y(t,n)-Y(t,1)}{n-1}] + \text{residual}$ where n and t are measured in years. Panel C shows the coefficents from the regressions $[Y(t+1,n) - Y(t,n)]^2 = \text{const} + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5) - Y(t,1)] + \phi_n(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] + \text{residual}$ where t is measured in months, n in years, and Y in percent. 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 $***$. *Source*: Fama and Bliss (1987) monthly observations from 1952:6 to 2004:12.

n	1	2	3	4	5
Panel A: Unconditional moments of the yield curve					
Mean	5.60	5.81	5.98	6.11	6.19
Yield change volatility	49.3	43.2	40.1	38.8	36.2
Skewness	0.83	0.79	0.78	0.77	0.77
Excess kurtosis	0.77	0.57	0.51	0.44	0.35
Panel B: Campbell–Shiller					
ϕ_n		-0.775^{**}	-1.1311^{***}	-1.5198^{***}	-1.4941^{***}
Standard error		(0.546)	(0.637)	(0.683)	(0.745)
Panel C: Volatility regression					
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)

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

Brandt and Chapman (2003), Piazzesi (2010), and Jacobs and Karoui (2009) document that yield volatility is time-varying and positively correlated with interest rates. To see this, Panel C in Table 1 shows the results from regressing squared monthly yield changes on the level, slope, and curvature of the yield curve — the three components identified by Litterman and Scheinkman (1991) that explain most return variability across the maturity spectrum. The level factor is strongly significant across all maturities confirming that yield volatility is related to the level of yields. 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 if the risk premium depends on volatility a positive relationship is expected. As discussed in Sec. 5.2, however, 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)).

The yield curve is upward sloping on average

A basic feature of the US term structure is that it is upward sloping on average: Panel A in Table 1 shows that for our sample period the five-year yield is 59 basis points higher than the one-year yield. The five-year yield is higher than the one-year yield in 79% of the months.

Volatility of yield changes is downward sloping on average

In contrast to the unconditional mean of yields, the unconditional volatility — defined as standard deviation — of yield changes is decreasing with maturity. Panel A in Table 1 shows that the monthly volatility decreases from 49.3 basis points for the one-year yield to 36.2 basis points for the five-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 is 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.

The distributions of yields are skewed and leptokurtic

While the first two moments of yields have received considerable attention in the literature, higher order moments are less studied.[3] In Panel A in Table 1 we see that the distributions of yields are skewed and leptokurtic since skewness and excess kurtosis are positive for all maturities.

3. Affine Term Structure Models

In this section, I describe affine term structure models. I first describe the dynamics of the short rate under the risk neutral measure and thereafter risk premium specifications. Different specifications lead to different model classes and their implications for the risk neutral dynamics are discussed.

[2] See Piazzesi (2005) for a detailed discussion.
[3] See Dutta and Babbel (2002) for a study of skewness and kurtosis in yields.

3.1. *Bond pricing*

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

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

and X_t follows an affine diffusion,

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

where W_t^Q 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. Dai and Singleton (2000) rank models according to the number of state variables entering the volatility matrix S_t and define an N-factor model with $m \leq N$ variables entering volatility as an $A_n(N)$ model. 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 a thorough description and discussion of restrictions for all three-factor models are given in Appendix A.

3.2. *Risk premia*

The stochastic discount factor M can be written as

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

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

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

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

$$S_t^{\frac{1}{2}}\Lambda_t = S_t\lambda_1$$

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

$$S_t^{\frac{1}{2}}\Lambda_t = S_t\lambda_1 + I^-\lambda_2 X_t, \tag{6}$$

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 *et al.* (2007) 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. Appendix A explains the parametrization of all three-factor essentially and extended affine models in detail.

To describe the connection between the Feller condition and the risk premium specification in a simple setting, we can look at the one-dimensional case. The short rate dynamics in the Cox–Ingersoll–Ross model which corresponds to the $A_1(1)$ model in Dai and Singleton (2000) is given as

$$dr_t = (\kappa_0 + \kappa_1 r_t)dt + \sigma\sqrt{r_t}dW_t^Q$$

under the risk neutral measure. The essentially and completely affine $A_1(3)$ risk premium is defined as

$$\Lambda_t = \lambda_2\sqrt{r_t}$$

and the dynamics of the short rate under the historical measure is

$$dr_t = (\kappa_0 + (\kappa_1 + \sigma\lambda_2)r_t)dt + \sigma\sqrt{r_t}dW_t^p.$$

We see that only the mean reversion coefficient in the drift is allowed to be different under the two measure. The extended affine $A_1(1)$ risk premium is

$$\Lambda_t = \frac{\lambda_1}{\sqrt{r_t}} + \lambda_2\sqrt{r_t}$$

and the historical short rate dynamics is

$$dr_t = ((\kappa_0 + \sigma\lambda_1) + (\kappa_1 + \sigma\lambda_2)r_t)dt + \sigma\sqrt{r_t}dW_t^p.$$

We see that both coefficients in the drift are allowed to be different under the two measures. However, the risk premium is not well defined if r_t is zero and the Feller condition which ensures that r_t stays away from zero needs to be imposed under both measures. In this example, the Feller condition is $\kappa_0 > \frac{\sigma^2}{2}$ under the risk neutral measure and $\kappa_0 + \sigma\lambda_1 > \frac{\sigma^2}{2}$ under the historical measure. Since the essentially affine model allows distributions under both measures for which the Feller condition is not satisfied, the extended model does not nest the essentially affine model. As we shall see later this restriction is binding under the risk neutral measure.

As illustrated in the one-factor example extended affine models has a more flexible risk premium specification but do not nest neither essentially nor completely affine models. Also, from a general equilibrium perspective the extended affine risk premium specification might be difficult to justify since agents in the economy become extremely averse to risk in periods where the risk is minimal. This is a consequence of the market price of risk being inversely related to the volatility of the state variables.

Finally, Duarte (2004) adds a constant term to the risk premium of essentially affine models given in Eq. (6) such that a non-linear term appears in the drift under P and names the models *semi-affine*. The risk premium is

$$S_t^{\frac{1}{2}}\Lambda_t = S_t^{\frac{1}{2}}\lambda_0 + S_t\lambda_1 + I^-\lambda_2 X_t, \tag{8}$$

where λ_0 is a $N \times 1$ vector. The extra term in semi-affine models potentially increases the flexibility of essentially affine models to better capture the historical distribution of yields while allowing strongly skewed distributions under the risk-neutral measure in contrast to extended models.

In contrast to extended models, semi-affine models nest the essentially affine models because the Feller condition is not imposed which in turn implies that semi-affine models are more flexible in generating a variety of distributions. Also, from a general equilibrium perspective the risk premium specification in semi-affine models seems more natural than in extended affine models since the price of risk is bounded as risk goes to zero in contrast to extended affine models where the price of risk explodes.

4. Estimation

In estimation, 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, every yield can be observed with error. Often, it is assumed that yields or a linear combination of yields are observed without error such that state variables can be extracted from yields.[5] Second, 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. Third, a non-linear drift of the state variables under the historical measure does not complicate the estimation. Finally, MCMC can easily handle parameter restrictions, while optimization algorithms of traditional frequentist methods often perform poorly in the presence of hard parameter constraints.[6]

4.1. *Estimation methodology*

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

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

[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 (2009). 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).

[5] See for example Dai and Singleton (2002), Duffee (2002), Cheridito *et al.* (2007), Almeida *et al.* (2011), Joslin *et al.* (2011), and Hamilton and Wu (2012).

[6] See for example Cheridito *et al.* (2007).

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

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

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; Besag, 1974) implies that samples are obtained from the target distribution by sampling from the full conditionals

$$p(K_0^Q \mid K_1^Q, \beta, \lambda, \delta, D, X, Y)$$
$$p(K_1^Q \mid K_0^Q, \beta, \lambda, \delta, D, X, Y)$$
$$\vdots$$
$$p(X \mid K_0^Q, K_1^Q, \beta, \lambda, \delta, D, Y)$$

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 $(K_0^Q, K_1^Q, \beta, \lambda, \delta, D, X)$ in the MCMC algorithm is obtained by drawing from the full conditionals

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

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.

4.1.1. *The conditionals $p(X \mid \Phi)$ and $p(Y \mid \Phi, X)$*

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

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

The continuous-time specification in (3) is approximated using an Euler scheme (letting $\Sigma = I_N$)[7]

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

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

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

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 $p(Y \mid \Phi, X)$ can be written as

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

4.1.2. *The hybrid MCMC algorithm*

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

$$p(\lambda \mid \Phi_{\backslash\lambda}, X, Y) \propto p(Y \mid \Phi, X) p(\lambda \mid \Phi_{\backslash\lambda}, X)$$
$$\propto p(X \mid \Phi) p(\lambda \mid \Phi_{\backslash\lambda}),$$

[7] 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 $\Phi_{\setminus\lambda}$ denotes the parameter vector without the parameter 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, non-informative prior, is adopted so $p(\lambda \,|\, \Phi_{\setminus\lambda}, X, Y) \propto p(X|\Phi)$ 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

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

since $p(X \,|\, \Phi)$ does not depend on D. σ^2 can therefore be Gibbs sampled from the inverse Wishart distribution, $\sigma^2 \sim IW(\sum_{t=1}^{T} \hat{e}'_t \hat{e}_t, kT)$.[8]

The conditional of the other model parameters is given as

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

which for none of the parameters K_0^Q, K_1^Q, β, and δ is a known distribution. To 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 normal distributed variable centered around Φ_j^i and accept it with probability $\min(1, \frac{f(\Phi_j^{i+1})}{f(\Phi_j^i)})$ where f is the density $p(\Phi_j \,|\, \Phi_{\setminus \Phi_j}, X, Y)$.

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

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

For $t = 0$ the conditional is

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

[8] Equivalent to an inverted gamma distribution.

while for $t = T$ the conditional is

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

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 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. I keep every 200th draw in the estimation period which leaves 20,000 draws, and parameter estimates are based on the mean of the draws. Further implementation details are given in Appendix B.1.

5. Results

Parameter estimates are given in Tables 2–7. I do not interpret individual parameters in the models because the main interest is on the economic implications of the models.

Table 2. Model estimates, essentially affine models (part 1). This table shows parameter estimates along with confidence bands for all three-factor essentially affine models.

	$A_0(3)$	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess
$K_1^Q(1,1)$	0.6250	0.0318	1.1323	0.2357
	(0.5344; 0.7075)	(0.0100; 0.0550)	(0.9843; 1.2964)	(0.1987; 0.2850)
$K_1^Q(1,2)$	0	0	−0.0770	−0.0461
			(−0.1439; −0.0182)	(−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	3.5617	−0.1634	−0.1661
	(4.0994; 5.2303)	(3.1709; 3.9545)	(−0.2978; −0.0301)	(−0.1884; −0.1385)
$K_1^Q(2,2)$	8.6864	0.0982	0.0549	0.1590
	(8.3999; 9.1701)	(0.0107; 0.1700)	(0.0222; 0.0978)	(0.1355; 0.1894)
$K_1^Q(2,3)$	0	4.0489	0	−0.0781
		(3.5131; 4.5243)		(−0.1988; −0.0011)
$K_1^Q(3,1)$	1.5609	1.9465	2.4236	−0.0033
	(0.9862; 1.8184)	(1.7062; 2.3039)	(2.1944; 2.7326)	(−0.0117; −0.0001)

Table 2. (*Continued*)

	$A_0(3)$	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess
$K_1^Q(3,2)$	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 3. Model estimates, essentially affine models (part 2). This table shows parameter estimates along with confidence bands for all three-factor essentially affine models. The parameters K_0^P and K_1^P are showed for completeness although they are functions of other parameters and are not estimated.

	$A_0(3)$ 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

Table 3. (*Continued*)

	$A_0(3)$ ess/ext	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess
$\lambda_1(3)$	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 4. Model estimates, extended affine models (part 1). This table shows parameter estimates along with confidence bands for all three-factor essentially affine models.

	$A_1(3)$ 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 5. Model estimates, extended affine models (part 2). This table shows parameter estimates along with confidence bands for all three-factor essentially affine models. The parameters K_0^P and K_1^P are showed for completeness although they are functions of other parameters and are not estimated.

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

Table 5. (*Continued*)

	$A_1(3)$ ext	$A_2(3)$ ext	$A_3(3)$ ext
$K_0^P(2)$	−0.0752	1.0119	1.0245
	(−0.7097; 0.5245)	(0.5188; 2.0401)	(0.5165; 2.0921)
$K_0^P(3)$	0.0364	0.7649	1.8007
	(−0.5922; 0.6914)	(−1.8154; 3.4713)	(0.5679; 3.7045)

Table 6. Model estimates, semi-affine models (part 1). This table shows parameter estimates along with confidence bands for all three-factor semi-affine models.

	$A_1(3)$ semi	$A_2(3)$ semi	$A_3(3)$ semi
$K_1^Q(1,1)$	0.0086	1.4202	0.2404
	(0.0004; 0.0225)	(1.1176; 1.7445)	(0.1973; 0.2910)
$K_1^Q(1,2)$	0	−0.2215	−0.0627
		(−0.3293; −0.1601)	(−0.1195; −0.0067)
$K_1^Q(1,3)$	0	0	−0.7770
			(−0.9932; −0.6401)
$K_1^Q(2,1)$	0.3224	−0.1222	−0.1361
	(−0.0881; 0.6494)	(−0.3498; −0.0044)	(−0.1779; −0.1035)
$K_1^Q(2,2)$	0.1286	0.0271	0.1824
	(0.0093; 0.2277)	(0.0045; 0.0610)	(0.1523; 0.2289)
$K_1^Q(2,3)$	6.2629	0	−0.0383
	(5.8798; 6.5914)		(−0.1285; −0.0011)
$K_1^Q(3,1)$	0.7957	3.7514	−0.0044
	(0.6289; 0.9521)	(3.4967; 4.0408)	(−0.0148; −0.0001)
$K_1^Q(3,2)$	−0.0618	−0.6462	−0.4413
	(−0.0760; −0.0514)	(−0.7262; −0.5616)	(−0.5704; −0.3645)
$K_1^Q(3,3)$	0.9978	0.7034	2.0676
	(0.8973; 1.1091)	(0.6085; 0.8105)	(1.8292; 2.3643)
$K_0^Q(1)$	0.2007	0.3151	0.4025
	(0.0512; 0.3598)	(0.0116; 0.9799)	(0.0150; 1.2020)
$K_0^Q(2)$	0	0.1265	0.0860
		(0.0061; 0.3487)	(0.0027; 0.2573)
$K_0^Q(3)$	0	0	0.1091
			(0.0030; 0.3853)
$\lambda_2(1,1)$	−0.2266	−0.5373	−0.1816
	(−0.4344; −0.0354)	(−1.1425; 0.0643)	(−0.4421; 0.0618)
$\lambda_2(1,2)$	0	−0.0120	0
		(−0.0656; 0.0432)	
$\lambda_2(1,3)$	0	0	0
$\lambda_2(2,1)$	44.521	1.7510	0
	(−8.8602; 102.24)	(0.1299; 3.5417)	

Table 6. (*Continued*)

	$A_1(3)$ semi	$A_2(3)$ semi	$A_3(3)$ semi
$\lambda_2(2,2)$	-0.7913	-0.2300	-0.1307
	$(-1.3310; -0.3108)$	$(-0.4277; -0.0436)$	$(-0.3832; 0.0754)$
$\lambda_2(2,3)$	9.5370	0	0
	(3.8843; 15.659)		
$\lambda_2(3,1)$	-4.2775	-0.0038	0
	$(-8.3133; -0.4301)$	$(-1.8916; 1.7677)$	
$\lambda_2(3,2)$	0.0344	-0.5012	0
	(0.0012; 0.0692)	$(-1.2073; 0.1956)$	
$\lambda_2(3,3)$	-0.3247	0.0734	0.0258
	$(-0.7180; 0.0642)$	$(-0.1973; 0.3397)$	$(-0.6131; 0.6924)$
σ^2	$4.87e-7$	$4.93e-7$	$5.05e-7$
	$(4.55e-7; 5.21e-7)$	$(4.61e-7; 5.27e-7)$	$(4.72e-7; 5.40e-7)$

Table 7. Model estimates, semi-affine models (part 2). This table shows parameter estimates along with confidence bands for all three-factor semi-affine models. The parameters K_1^P are showed for completeness although they are functions of other parameters and are not estimated.

	$A_1(3)$ semi	$A_2(3)$ semi	$A_3(3)$ semi
$\lambda_1(1)$	0	0	0
$\lambda_1(2)$	354.7	0	0
	$(-262.2; 987.4)$		
$\lambda_1(3)$	-39.59	0.4730	0
	$(-85.59; 5.146)$	$(-2.7826; 3.7789)$	
$\lambda_0(1)$	1.1072	1.7070	1.4964
	(0.2389; 2.0524)	(0.4543; 2.9491)	(0.2609; 2.8186)
$\lambda_0(2)$	-2.6947	1.1442	0.2122
	$(-6.1536; 0.7308)$	(0.2694; 2.0719)	$(-0.5954; 1.1719)$
$\lambda_0(3)$	3.2914	1.9358	0.1543
	$(-0.1475; 6.7605)$	$(-1.5803; 5.6885)$	$(-1.0367; 1.3608)$
δ_0	0.0176	0.0215	0.0070
	(0.0174; 0.0178)	(0.0206; 0.0224)	(0.0058; 0.0092)
$\delta_x(1)$	0.0024	0.0052	-0.0010
	(0.0023; 0.0026)	(0.0040; 0.0063)	$(-0.0015; -0.0007)$
$\delta_x(2)$	$5.63e-6$	0.0012	0.0024
	$(1.14e-6; 11.1e-6)$	(0.0010; 0.0015)	(0.0019; 0.0028)
$\delta_x(3)$	4.10e−4	0.0023	0.0120
	$(3.45e-4; 5.03e-4)$	(0.0018; 0.0028)	(0.0100; 0.0137)
$\beta_2(1)$	12298	0	0
	(8798; 15703)		
$\beta_3(1)$	64.67	0.7379	0
	(43.93; 92.34)	(0.0155; 3.1019)	
$\beta_3(2)$	0	2.3153	0
		(1.4752; 3.5721)	

Table 7. (*Continued*)

	$A_1(3)$ semi	$A_2(3)$ semi	$A_3(3)$ semi
$K_1^P(1,1)$	0.2352	1.9576	0.4220
	(0.0432; 0.4417)	(1.3053; 2.6334)	(0.1822; 0.6797)
$K_1^P(1,2)$	0	−0.2215	−0.0627
		(−0.3293; −0.1601)	(−0.1195; −0.0067)
$K_1^P(1,3)$	0	0	−0.7770
			(−0.9932; −0.6401)
$K_1^P(2,1)$	−44.20	−0.1222	−0.1361
	(−101.89; 9.159)	(−0.3498; −0.0044)	(−0.1779; −0.1035)
$K_1^P(2,2)$	0.9199	0.2571	0.3132
	(0.4277; 1.4603)	(0.0701; 0.4554)	(0.1073; 0.5641)
$K_1^P(2,3)$	−3.2741	0	−0.0383
	(−9.4159; 2.3856)		(−0.1285; −0.0011)
$K_1^P(3,1)$	5.0731	3.7552	−0.0044
	(1.2039; 9.1203)	(1.9899; 5.6890)	(−0.0148; −0.0001)
$K_1^P(3,2)$	−0.0961	−0.1450	−0.4413
	(−0.1353; −0.0618)	(−0.8622; 0.5617)	(−0.5704; −0.3645)
$K_1^P(3,3)$	1.3224	0.6300	2.0418
	(0.9367; 1.7109)	(0.3427; 0.9294)	(1.4177; 2.6813)

5.1. *Campbell–Shiller regression coefficients*

Table 8 shows the Campbell–Shiller regression coefficients for the models.[9]

We see that the ability of essentially affine models to capture the CS regression coefficients decrease in the number of volatility factors. For example, the coefficients for the five-year bond are −0.591, 0.314, 0.690, and 1.309 for the models with 0, 1, 2, and 3 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, and all models with stochastic volatility miss the slope and sign of the CS coefficients. This is consistent with findings in Dai and Singleton (2002).

We see that the results for the semi-affine models are comparable to those of the essentially affine models, so the non-linear term in semi-affine models does not help in replicating time-varying risk premia. In contrast, while extended affine models are not able to beat the $A_0(3)$ model in capturing the CS coefficients, they do better than their essentially affine counterparts. We see a

[9] 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. The regression coefficients are calculated from simulated data and details about the simulation procedure is given in Appendix B.2.

Table 8. Model-implied Campbell–Shiller regression coefficients. This table shows the regression coefficients from the regressions $Y(t+1, n-1) - Y(t, n) = \text{const} + \phi_n[\frac{Y(t,n)-Y(t,1)}{n-1}]+$ residual where n and t are measured in years. The procedure for calculating the coefficients in the models takes into account finite-sample bias by simulating as explained in the text.

n	2	3	4	5
Actual	−0.775	−1.131	−1.520	−1.494
$A_0(3)$ ess	−0.206	−0.316	−0.447	−0.591
	(−1.089; 0.497)	(−1.338; 0.519)	(−1.561; 0.479)	(−1.798; 0.417)
$A_1(3)$ ess	0.302	0.243	0.262	0.314
	(−0.364; 0.887)	(−0.425; 0.892)	(−0.430; 0.971)	(−0.418; 1.075)
$A_2(3)$ ess	0.558	0.587	0.636	0.690
	(−0.087; 1.169)	(−0.124; 1.231)	(−0.110; 1.338)	(−0.084; 1.415)
$A_3(3)$ ess	1.080	1.158	1.236	1.309
	(0.796; 1.323)	(0.827; 1.424)	(0.870; 1.530)	(0.946; 1.602)
$A_1(3)$ ext	0.231	0.089	0.071	0.134
	(−0.293; 0.649)	(−0.443; 0.541)	(−0.502; 0.532)	(−0.441; 0.598)
$A_2(3)$ ext	0.381	0.340	0.337	0.359
	(−0.138; 0.872)	(−0.232; 0.878)	(−0.289; 0.924)	(−0.302; 0.981)
$A_3(3)$ ext	0.282	0.233	0.192	0.147
	(−0.184; 0.765)	(−0.326; 0.826)	(−0.421; 0.851)	(−0.502; 0.846)
$A_1(3)$ semi	0.422	0.358	0.379	0.434
	(−0.155; 0.960)	(−0.268; 0.944)	(−0.255; 0.992)	(−0.204; 1.061)
$A_2(3)$ semi	0.619	0.618	0.636	0.662
	(−0.068; 1.115)	(−0.093; 1.133)	(−0.087; 1.165)	(−0.070; 1.188)
$A_3(3)$ semi	1.292	1.401	1.476	1.530
	(0.857; 1.512)	(0.909; 1.658)	(0.921; 1.760)	(0.938; 1.858)

slight improvement in time-varying predictability in the $A_1(3)$ model, a sizeable improvement in the $A_2(3)$ model and a dramatic improvement in the $A_3(3)$ model.

In the literature 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 semi-affine models inherit this inability. In contrast, the results in this paper show that in terms of predictability the extended affine with two or three volatility factors do as well as their one volatility factor counterparts.

5.2. *Conditional volatility*

In Sec. 2, monthly squared yield changes are regressed 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 the significant level coefficients along with the slope and curvature coefficients. Table 9 shows the model-implied regression coefficients.

Table 9. Volatility regression. This table shows for the affine models the coefficents from the regressions $[Y(t+1,n) - Y(t,n)]^2 = \text{const} + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5) - Y(t,1)] + \phi_n(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] +$ residual where t is measured in months, n in years, and Y in percent. The regression coefficients are calculated using simulated yields (in percent).

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)

Table 9. (*Continued*)

n	1	2	3	4	5
$A_3(3)$ 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)

Table 9. (*Continued*)

n	1	2	3	4	5
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)
$A_1(3)$ semi					
Level	0.041	0.034	0.030	0.027	0.024
	(0.036; 0.045)	(0.030; 0.038)	(0.027; 0.033)	(0.024; 0.030)	(0.022; 0.027)
Slope	0.046	0.037	0.033	0.029	0.027
	(0.034; 0.058)	(0.028; 0.047)	(0.025; 0.041)	(0.022; 0.037)	(0.020; 0.034)
Curvature	0.079	0.063	0.055	0.049	0.044
	(0.058; 0.104)	(0.045; 0.084)	(0.038; 0.074)	(0.034; 0.067)	(0.031; 0.060)
$A_2(3)$ semi					
Level	0.042	0.034	0.030	0.026	0.024
	(0.037; 0.047)	(0.031; 0.038)	(0.027; 0.033)	(0.024; 0.029)	(0.022; 0.027)
Slope	0.056	0.043	0.037	0.032	0.029
	(0.042; 0.073)	(0.032; 0.054)	(0.028; 0.046)	(0.024; 0.040)	(0.022; 0.036)
Curvature	0.085	0.061	0.052	0.045	0.039
	(0.053; 0.132)	(0.061; 0.089)	(0.052; 0.074)	(0.045; 0.063)	(0.029; 0.056)
$A_3(3)$ semi					
Level	0.042	0.033	0.028	0.025	0.022
	(0.038; 0.046)	(0.030; 0.036)	(0.025; 0.031)	(0.022; 0.027)	(0.020; 0.025)
Slope	0.014	−0.006	−0.010	−0.010	−0.009
	(−0.005; 0.034)	(−0.019; 0.008)	(−0.020; 0.001)	(−0.019; −0.001)	(−0.017; −0.001)
Curvature	0.084	0.002	−0.016	−0.020	−0.019
	(0.048; 0.119)	(−0.024; 0.025)	(−0.038; 0.003)	(−0.040; −0.002)	(−0.036; −0.003)

Since the $A_0(3)$ model does not accommodate stochastic volatility, coefficients for every maturity are estimated to be zero which is not consistent with the data. All models with stochastic volatility capture the positive sign and the downward sloping curve with respect to maturity of the level coefficients. The models largely agree on the size of the level coefficients although the extended $A_1(3)$ model estimates the coefficients somewhat lower than other stochastic volatility models. However, 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 five-year coefficient is 0.044 while in the affine models with stochastic volatility it is estimated to be in the range 0.018–0.027 and this difference is statistically significant.

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 five-year maturity) and the curvature coefficients are positive, the models with one and two volatility factors predict positive coefficients on slope and curvature while the models with three stochastic factors predict negative coefficients for long maturities and positive coefficients for short maturities.[10]

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 9 as the most important factor in the difference of fit between the two model classes.

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 1 graphs the model-implied and EGARCH(1,1) conditional volatility for the five-year yield.[11] In

[10] In a previous version of the paper, the robustness of the results are tested by running a volatility regression 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 and available on request.

[11] 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 $t = 1, \ldots, 631$ conditional volatility is estimated on basis of the 20,000 draws of time t conditional volatility.

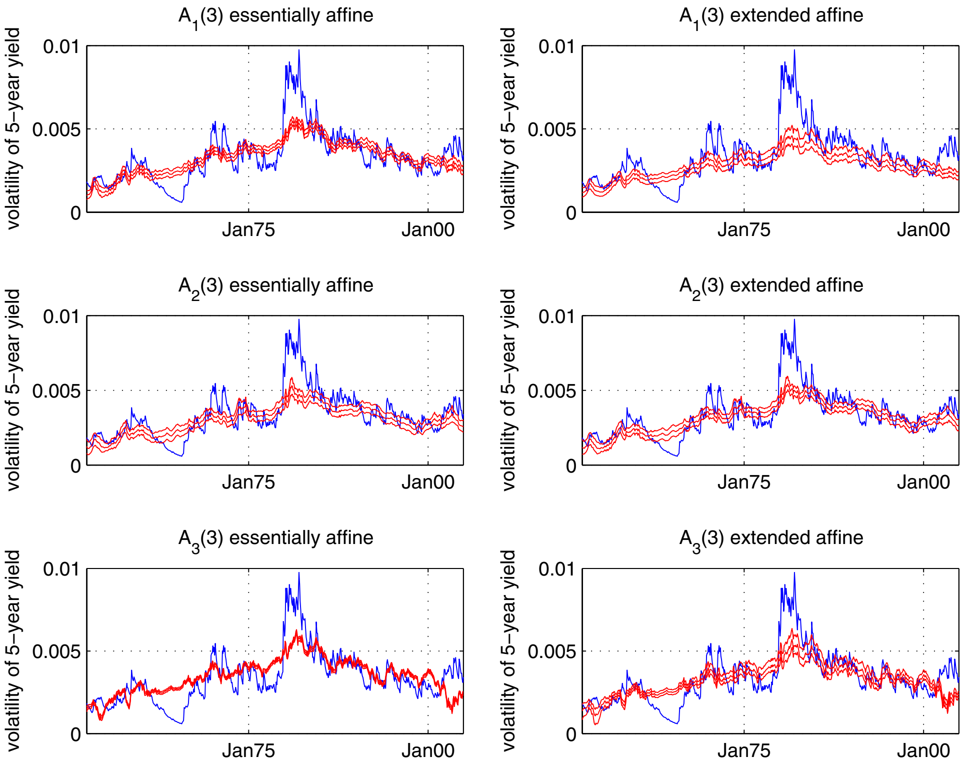

Fig. 1. Actual and model-implied five-year conditional volatility. This figure depicts for the affine models with stochastic volatility the estimate and confidence band of the monthly conditional volatility in the five-year yield along with an EGARCH(1,1) estimate.

the following we focus on the essentially and extended affine models for brevity, but results for the semi-affine models are similar.

The figure shows that all models to some extent capture the persistence in conditional volatility apart from the period in the beginning of the 80s. 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 all models fail to capture the high volatility during the Fed experiment from October 1979 to October 1982. Volatility in the five-year yield across the models is roughly 50% lower than actual volatility. To investigate the influence of the Fed experiment on the volatility regression results, Table 10 shows the regression coefficients for the period before and after the Fed experiment.

Compared to the coefficients obtained using the whole sample the results are strikingly different: the level coefficient is only about half the size before the Fed experiment and about one-third after the Fed experiment. In both

Table 10. Volatility regression in subperiods. This table shows the coefficients from the same volatility regression as Table 1 except that the regression is split into the period before the Fed Experiment and after the Fed Experiment. The regression is $[Y(t+1,n) - Y(t,n)]^2 = \text{const} + \phi_n(1)[Y(t,5)] + \phi_n(2)[Y(t,5) - Y(t,1)] + \phi_n(3)[Y(t,5) + Y(t,1) - 2Y(t,3)] + \text{residual}$ 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 $***$.

n	1	2	3	4	5
	Whole sample period 1952:6 to 2004:12				
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				
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				
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)

subperiods the level effect is still statistically 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) and 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. 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 11 shows that the correlations between EGARCH and model-implied volatility are positive and in the range of 67.5% to 80.8% across models and maturity for the whole sample and similar correlations are found in the

Table 11. Correlation between model-implied and actual conditional volatility. This table shows the correlation between model-implied monthly conditional volatility and an EGARCH (1,1) estimate of monthly conditional volatility.

	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess	$A_1(3)$ ext	$A_2(3)$ ext	$A_3(3)$ ext
	Whole sample period 1952:6 to 2004:12					
One-year	69.9	67.5	73.1	72.5	68.8	72.9
	(68.8; 70.9)	(63.9; 70.1)	(72.8; 73.6)	(71.5; 73.6)	(65.4; 71.1)	(71.4; 74.0)
Two-year	74.0	71.4	72.3	75.1	72.8	73.8
	(73.2; 74.7)	(67.2; 74.4)	(72.0; 72.8)	(74.3; 75.9)	(69.1; 74.8)	(73.1; 74.7)
Three-year	78.1	76.6	74.3	78.6	77.9	76.1
	(77.4; 78.7)	(73.1; 79.0)	(73.9; 75.0)	(77.9; 79.4)	(75.4; 79.5)	(74.8; 77.2)
Four-year	80.7	79.2	76.1	80.8	80.5	77.9
	(79.9; 81.3)	(75.8; 81.5)	(75.7; 76.7)	(80.1; 81.6)	(78.1; 81.9)	(76.4; 79.2)
Five-year	78.7	78.5	73.1	78.6	79.4	74.8
	(77.9; 79.5)	(75.9; 80.3)	(72.8; 73.7)	(77.7; 79.5)	(77.8; 80.8)	(73.4; 76.2)
Average	76.3	74.6	73.8	77.2	75.9	75.1
	(75.6; 76.9)	(71.3; 76.9)	(73.5; 74.3)	(76.4; 77.9)	(73.2; 77.5)	(74.2; 76.0)
	Period before the Fed experiment 1952:6 to 1979:9					
One-year	77.2	73.0	71.9	76.7	74.2	75.4
	(75.6; 78.6)	(66.2; 79.1)	(70.9; 72.7)	(75.1; 78.2)	(65.9; 79.6)	(74.1; 76.9)
Two-year	72.3	66.6	67.2	70.4	69.3	70.3
	(70.4; 74.0)	(58.6; 74.1)	(66.6; 67.9)	(68.2; 72.3)	(59.9; 75.3)	(68.7; 71.9)
Three-year	75.5	72.0	71.5	74.0	74.9	73.7
	(73.7; 77.1)	(65.6; 77.1)	(71.0; 72.2)	(72.0; 75.7)	(68.3; 79.1)	(72.3; 74.8)
Four-year	76.9	74.4	72.8	75.3	77.2	74.8
	(75.2; 78.4)	(68.7; 78.8)	(72.2; 73.5)	(73.5; 77.0)	(71.8; 80.7)	(73.4; 76.0)

Table 11. (*Continued*)

	$A_1(3)$ ess	$A_2(3)$ ess	$A_3(3)$ ess	$A_1(3)$ ext	$A_2(3)$ ext	$A_3(3)$ ext
Five-year	67.3	67.4	62.5	65.8	69.3	64.1
	(65.4; 69.1)	(62.9; 71.1)	(62.0; 63.1)	(63.8; 67.7)	(65.4; 73.0)	(63.0; 65.3)
Average	73.8	70.7	69.2	72.4	73.0	71.7
	(72.1; 75.4)	(64.5; 75.7)	(68.7; 69.8)	(70.6; 74.1)	(66.4; 77.2)	(70.4; 72.8)
Period after the Fed experiment 1982:11 to 2004:12						
One-year	79.8	62.2	76.3	80.7	65.3	79.7
	(78.0; 79.9)	(46.7; 78.1)	(75.3; 77.2)	(79.4; 82.0)	(48.6; 77.4)	(78.1; 81.2)
Two-year	74.4	66.1	63.0	73.8	70.2	67.6
	(72.8; 76.0)	(54.0; 76.0)	(62.2; 64.5)	(72.1; 75.4)	(58.4; 76.4)	(64.1; 71.2)
Three-year	70.9	65.4	57.2	69.8	69.5	61.7
	(68.9; 72.8)	(55.0; 73.0)	(56.3; 58.8)	(67.9; 71.8)	(60.7; 74.2)	(57.9; 65.5)
Four-year	70.6	66.7	56.7	59.5	70.2	60.5
	(68.6; 72.4)	(57.9; 72.9)	(55.8; 58.2)	(67.6; 71.6)	(63.4; 74.0)	(57.1; 64.2)
Five-year	69.3	67.4	54.6	67.9	70.4	57.9
	(67.2; 71.3)	(60.1; 72.3)	(53.8; 55.9)	(65.7; 70.1)	(65.2; 73.6)	(54.8; 61.5)
Average	73.0	65.5	61.6	72.3	69.1	65.5
	(71.4; 74.6)	(55.0; 74.2)	(60.9; 62.9)	(70.7; 74.0)	(59.5; 74.6)	(62.5; 68.6)

subperiods before and after the Fed experiment although the latter period has somewhat smaller correlations. This result is consistent with Almeida *et al.* (2006) who find similar positive correlations between conditional volatilities of yields of different maturities and GARCH estimates, while Collin-Dufresne *et al.* (2006) 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.[12]

Comparing essentially and extended models, the correlations suggest that extended $A_2(3)$ and $A_3(3)$ models do slightly better in capturing conditional volatility than their essentially affine counterparts since they have higher average correlations in the whole period and in both subsamples. Ahn *et al.* (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 matching time-varying volatility in moving from

[12] See Joslin (2006) and Jacobs and Karoui (2009) for an elaboration on this point.

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.

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. However, none of the models can capture the high volatility during the monetary experiment.

Collin-Dufresne *et al.* (2009) show that affine models with so-called Unspanned Stochastic Volatility (USV) have better volatility dynamics, but Creal and Wu (2015) show that USV models cannot match the high volatility of the early 1980s either. It seems that one has to move outside the pure affine models to capture the volatility dynamics over long periods of time, such as regime-switching models (Bansal *et al.*, 2004; Dai *et al.*, 2007) or non-linear models (Feldhütter *et al.*, 2015).

5.3. *Unconditional moments*

Turning to unconditional moments, Tables 12 and 13 show model-implied unconditional mean and volatility of yields.

All models regardless of risk premium specification and number of volatility factors capture the size and slope of the mean and volatility curves.

Table 12. 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	5.60	5.81	5.98	6.11	6.19
Panel A					
$A_0(3)$ 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)

Table 12. (*Continued*)

n	1	2	3	4	5
$A_1(3)$ 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)
Panel B					
$A_1(3)$ 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)
Panel C					
$A_1(3)$ semi					
Mean	7.13	7.38	7.56	7.68	7.75
Median	6.06	6.31	6.53	6.68	6.75
	(3.62; 13.32)	(3.86; 13.10)	(4.03; 13.43)	(4.13; 13.60)	(4.23; 13.66)
$A_2(3)$ semi					
Mean	6.44	6.70	6.91	7.05	7.15
Median	6.16	6.38	6.60	6.73	6.80
	(4.26; 11.32)	(4.44; 11.27)	(4.60; 11.72)	(4.74; 11.78)	(4.86; 11.92)
$A_3(3)$ semi					
Mean	6.91	7.13	7.31	7.43	7.51
Median	6.01	6.25	6.44	6.59	6.69
	(3.25; 17.39)	(3.41; 17.69)	(3.53; 17.65)	(3.61; 17.57)	(3.70; 17.31)

However, the unconditional mean of yields is hard to pin down with reasonable precision. For example, the average five-year yield is 6.19% while the 95% confidence band in the essentially affine $A_1(3)$ model is [2.84%, 40.90%]. Thus, the average yield curve is not a moment that is useful to statistically discriminate between models.

Table 13. Unconditional volatility of yields. The first line in this table shows the unconditional volatility (standard deviation) in basis points of monthly yield changes in the data where n denotes maturity. The next line 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	49.3	43.2	40.1	38.8	36.2
Panel A					
$A_0(3)$ 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)
Panel B					
$A_1(3)$ 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)
Panel C					
$A_1(3)$ semi					
Mean	47.5	43.2	40.6	38.5	36.7
Median	45.8	41.6	39.1	37.1	35.4
	(31.0; 70.0)	(28.0; 63.0)	(26.4; 59.3)	(25.0; 56.1)	(23.8; 53.2)
$A_2(3)$ semi					
Mean	47.7	43.0	39.9	37.6	35.8
Median	46.6	41.8	38.9	36.7	35.0
	(37.0; 66.2)	(33.5; 59.7)	(31.1; 55.5)	(29.2; 52.3)	(27.9; 50.1)

Table 13. (*Continued*)

n	1	2	3	4	5
$A_3(3)$ semi					
Mean	50.7	44.3	41.9	39.3	37.2
Median	48.5	43.3	40.2	37.7	35.6
	(34.0; 80.0)	(30.7; 72.4)	(28.2; 67.7)	(26.3; 63.7)	(24.7; 60.3)

Extended and semi-affine models estimate volatility with much less uncertainty than the essentially affine models. For example, the length of the confidence band for the volatility of the five-year yield in the essentially affine $A_1(3)$ model is 88.2 basis points while it is 21.0 in the extended $A_1(3)$ model and 29.4 in the semi-affine $A_1(3)$ model. The $A_0(3)$ model which is relieved from the task of fitting conditional volatility estimates the volatility of the five-year yield with a confidence band of only 3.7 basis points.

Differences in the models become clear when we examine higher order moments. Tables 14 and 15 report the skewness and (excess) kurtosis of yields. Since yields in the $A_0(3)$ model are normal distributed, skewness and kurtosis are zero for all yields and the results are therefore not shown for this model.

Table 14. Unconditional skewness of yields. The first line in this table shows the unconditional skewness of monthly yields in the data where n denotes maturity. The next line show the model-implied unconditional mean, median, and confidence bands of unconditional skewness for the estimated models. These are calculated on basis of simulated yields as explained in the text.

n	1	2	3	4	5
Actual	0.83	0.79	0.78	0.77	0.77
Panel A					
$A_1(3)$ ess					
Mean	2.16	2.21	2.25	2.26	2.27
Median	2.13	2.18	2.21	2.22	2.23
	(1.26; 3.19)	(1.41; 3.21)	(1.47; 3.28)	(1.48; 3.35)	(1.50; 3.36)
$A_2(3)$ ess					
Mean	1.67	1.77	1.85	1.91	1.95
Median	1.68	1.73	1.79	1.83	1.84
	(−0.20; 3.41)	(0.23; 3.49)	(0.41; 3.71)	(0.64; 3.82)	(0.66; 3.88)
$A_3(3)$ ess					
Mean	1.55	1.55	1.54	1.53	1.52
Median	1.52	1.51	1.50	1.50	1.49
	(1.02; 2.28)	(1.01; 2.28)	(1.01; 2.27)	(1.00; 2.27)	(1.00; 2.27)

Table 14. (*Continued*)

n	1	2	3	4	5
Panel B					
$A_1(3)$ ext					
Mean	0.84	0.88	0.90	0.92	0.93
Median	0.81	0.84	0.86	0.88	0.89
	(0.55; 1.35)	(0.59; 1.38)	(0.61; 1.39)	(0.63; 1.39)	(0.65; 1.41)
$A_2(3)$ ext					
Mean	0.77	0.84	0.88	0.91	0.93
Median	0.78	0.85	0.89	0.91	0.93
	(0.32; 1.16)	(0.42; 1.18)	(0.53; 1.19)	(0.61; 1.21)	(0.65; 1.23)
$A_3(3)$ ext					
Mean	0.88	0.90	0.90	0.90	0.90
Median	0.87	0.89	0.90	0.89	0.89
	(0.62; 1.18)	(0.65; 1.19)	(0.65; 1.20)	(0.64; 1.20)	(0.64; 1.20)
Panel C					
$A_1(3)$ semi					
Mean	0.76	0.79	0.80	0.79	0.79
Median	0.66	0.69	0.71	0.71	0.70
	(0.37; 1.49)	(0.40; 1.52)	(0.46; 1.50)	(0.48; 1.49)	(0.48; 1.45)
$A_2(3)$ semi					
Mean	0.55	0.58	0.60	0.61	0.62
Median	0.50	0.53	0.55	0.57	0.57
	(0.29; 1.08)	(0.33; 1.14)	(0.37; 1.18)	(0.39; 1.21)	(0.40; 1.22)
$A_3(3)$ semi					
Mean	0.87	0.86	0.84	0.82	0.81
Median	0.80	0.79	0.78	0.76	0.74
	(0.54; 1.75)	(0.53; 1.71)	(0.52; 1.67)	(0.50; 1.64)	(0.49; 1.61)

Table 15. Unconditional excess kurtosis of yields. The first line in this table shows the unconditional excess kurtosis of monthly yields in the data where n denotes maturity. The next line show the model-implied unconditional mean, median, and confidence bands of unconditional skewness for the estimated models. These are calculated on basis of simulated yields as explained in the text.

n	1	2	3	4	5
Actual	0.77	0.57	0.51	0.44	0.35
Panel A					
$A_1(3)$ ess					
Mean	7.39	7.54	7.62	7.66	7.68
Median	6.60	6.82	6.98	6.97	7.00
	(2.43; 17.23)	(2.43; 17.35)	(2.50; 17.54)	(2.47; 17.45)	(2.42; 17.36)
$A_2(3)$ ess					
Mean	5.67	5.93	6.19	6.43	6.62
Median	4.46	4.53	4.66	4.78	4.80
	(0.80; 18.38)	(0.81; 20.17)	(0.84; 20.62)	(0.87; 20.85)	(0.95; 22.26)

Table 15. (*Continued*)

n	1	2	3	4	5
$A_3(3)$ ess					
Mean	3.58	3.54	3.50	3.46	3.44
Median	3.26	3.20	3.12	3.07	3.05
	(0.99; 7.55)	(0.97; 7.65)	(0.93; 7.72)	(0.91; 7.77)	(0.91; 7.81)
Panel B					
$A_1(3)$ ext					
Mean	1.22	1.28	1.31	1.34	1.36
Median	1.02	1.06	1.11	1.13	1.14
	(0.43; 2.99)	(0.47; 3.09)	(0.52; 3.20)	(0.51; 3.23)	(0.53; 3.24)
$A_2(3)$ ext					
Mean	1.17	1.26	1.32	1.37	1.40
Median	1.09	1.18	1.25	1.30	1.32
	(0.49; 2.20)	(0.56; 2.32)	(0.62; 2.44)	(0.65; 2.56)	(0.66; 2.58)
$A_3(3)$ ext					
Mean	1.20	1.23	1.23	1.23	1.23
Median	1.12	1.15	1.15	1.15	1.14
	(0.54; 2.24)	(0.53; 2.30)	(0.52; 2.34)	(0.51; 2.36)	(0.50; 2.37)
Panel C					
$A_1(3)$ semi					
Mean	1.05	1.08	1.06	1.03	0.99
Median	0.65	0.69	0.69	0.67	0.65
	(0.21; 3.18)	(0.23; 3.31)	(0.26; 3.24)	(0.26; 3.10)	(0.25; 2.93)
$A_2(3)$ semi					
Mean	0.50	0.53	0.55	0.56	0.57
Median	0.36	0.38	0.41	0.43	0.44
	(0.12; 1.65)	(0.13; 1.67)	(0.14; 1.79)	(0.15; 1.85)	(0.15; 1.91)
$A_3(3)$ semi					
Mean	1.23	1.19	1.15	1.11	1.07
Median	0.89	0.85	0.83	0.79	0.76
	(0.35; 4.56)	(0.34; 4.41)	(0.31; 4.28)	(0.30; 4.21)	(0.28; 4.08)

It is clear from the tables that the distribution of yields in essentially affine models are too skewed and leptokurtic compared to what we see in the data. For example, the skewness and kurtosis of the five-year yield in the essentially affine $A_1(3)$ model are 2.27 and 7.68 while they are 0.77 and 0.35 in the actual data and the differences are statistically highly significant. The skewness and kurtosis decrease with the number of volatility factors in the essentially affine models, but even in the $A_3(3)$ model they are larger than in the data and the difference remains statistically significant. In contrast, extended and semi-affine models do better in capturing the skewness and kurtosis in yields. The models estimate the skewness of yields to be close to the skewness in the actual data and the differences are statistically insignificant. Also model-implied kurtosis is closer to the actual values. For example, the actual kurtosis in the five-year yield is 0.35 while the essentially affine models estimate it to be in the range 3.44–7.68 and the extended models estimate it to be in the range 1.23–1.40. For long maturities the difference between actual and model-implied kurtosis remains statistically significant in the extended models. In contrast semi-affine models have kurtosis coefficients close to those in the data.

To help shed light on the reason why the distributions of yields are so skewed in essentially affine models it is useful to compare essentially and extended $A_1(3)$ models. The $A_1(3)$ models are chosen for comparison since they are the most commonly used three-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 of the bond with longest maturity — in this case the five-year yield — and has a low mean-reversion under the risk-neutral measure while the other two factors have 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 low mean-reversion but this would limit the model's ability to fit a wide variety of term-structure shapes.[13] Since the two non-volatility factors "die out" rather quickly the five-year yield is close to being modeled as an affine function of a one-factor CIR process.[14] As an approximation it is therefore reasonable to assume that the five-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 absolute value of skewness are unchanged by an affine transformation the five-year yield Y^5 has skewness $\text{sign}(\delta_x(1)) \times \sqrt{\frac{2}{K_0(1)}}$ and kurtosis $\frac{3}{K_0(1)}$. While all the parameters $\delta_0, \delta_x(1), K_0(1)$, and

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

[14] If one of the Gaussian factors exhibits low mean reversion instead of the CIR process, the distribution of the five-year yield would be close to Gaussian and the skewness and excess kurtosis of the five-year yield therefore be close to zero under both the pricing and historical measure, which provides minimal flexibility in generating skewed and leptokurtic distributions.

$K_1(1,1)$ determine the mean of X_t and three of the parameters determine the variance of X_t the only parameter determining kurtosis and absolute value of skewness is $K_0(1)$. In the $A_1(3)$ essentially affine model, the estimate of $K_0^P(1)$ is 0.3741. According to the previous argument $K_0^P(1) = 0.3741$ implies a skewness and kurtosis of $\sqrt{\frac{2}{0.3741}} = 2.31$ and $\frac{3}{0.3741} = 8.02$ which is close to the model-implied skewness and kurtosis of 2.27 and 7.68. In the $A_1(3)$ extended affine model we have $K_0^P(1) = 2.2731$ and while $\sqrt{\frac{2}{2.2731}} = 0.94$ and $\frac{3}{2.2731} = 1.32$ we have that model-implied skewness and kurtosis are 0.93 and 1.36. In both cases, approximating the dynamics of the five-year yield as a one-factor CIR process implies skewness and kurtosis that are close to the model-implied skewness and kurtosis. From Tables 14 and 15 we also see that the parameter $K_0^P(1)$ not only determines the third and fourth moments of the five-year yield but also plays an important role in determining the higher order moments of the shorter-maturity yields, since the moments of the shorter-maturity yields are similar to those of the five-year yield.

When the Q and P dynamics share parameters the Q dynamics tends to dominate in determining the parameters in estimation and therefore in the essentially affine $A_1(3)$ model the skewness and kurtosis of yields under P is primarily determined by the skewness and kurtosis under Q since the Q and P dynamics share the parameter $K_0(1)$.[15] 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 skewness and kurtosis under P such that model-implied third and fourth moments of yields more closely resembles historical third and fourth moments. To put it simple, the actual distribution of yields in the essentially affine $A_1(3)$ model inherits a counterfactual skewed and leptokurtic distribution from the risk-neutral dynamics while the extra risk premium parameter in the extended affine $A_1(3)$ model allows 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 leptokurtic distributions. The Feller condition requires $K_0(1) > 0.5$ and therefore skewness and excess

[15] An indication that the cross section of yields 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.

kurtosis in the distribution of the five-year yield cannot be higher than 2 and 6 in the $A_1(3)$ model. 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 is binding under the risk-neutral measure since the estimate of $K_0^Q(1)$ is less than 0.5 in the essentially affine $A_1(3)$ model.

The Feller restriction sets a tight upper limit on the skewness and kurtosis of unconditional distribution of yields and this restriction is even tighter when looking at the conditional distributions. As shown in Appendix C kurtosis and the absolute value of skewness of the conditional distribution of a CIR process $Y_{t+\tau|t}$ are monotone increasing in τ, go to zero as τ goes to zero, and go to excess kurtosis and the absolute value of skewness of the unconditional distribution as $\tau \to \infty$. Therefore, skewness and kurtosis of conditional distributions of yields are bounded to be lower than skewness and kurtosis of unconditional distributions.

Figure 2 shows the historical distribution of the five-year yield along with model-implied distributions in the models.[16] The figure clearly shows that the actual distribution of the five-year yield is matched much better by extended and semi-affine models — consistent with their improved ability to capture higher order moments.

The extended affine risk premium specification allows increased flexibility in two directions compared to the essentially affine. First, it allows a risk premium, λ_1, on the constant term in the drift of the CIR processes. Second, it allows an increased flexibility in the risk premium on the mean reversion matrix, λ_2, since there are risk premia on the correlations between the CIR processes.[17] As shown in the previous section the extra flexibility in λ_1 in extended models is important for allowing differences in shapes of the distributions of yields, but the cost of the flexibility is that the Feller condition is required to hold. In addition, the added flexibility in λ_1 does not help in replicating time-varying risk premia. In estimation of hybrid essentially/extended affine models where the specification of λ_1 is that of extended models and the specification of λ_2 is that of essentially affine models, the resulting Campbell–Shiller regression coefficients are close to those of the essentially affine models. Therefore, the added flexibility of λ_2 in extended models is important for the ability to generate time-varying risk premia. This added flexibility allows that the loading of other volatility factors change

[16] Model-implied densities in the figure are calculated by estimating the cumulative distribution function of the five-year yield for a fine grid of values. The densities are therefore based on all the draws in the MCMC sampler and represent the "average density".

[17] Equations (A.1)–(A.3) in Appendix A illustrate the increased flexibility on λ_1 and λ_2.

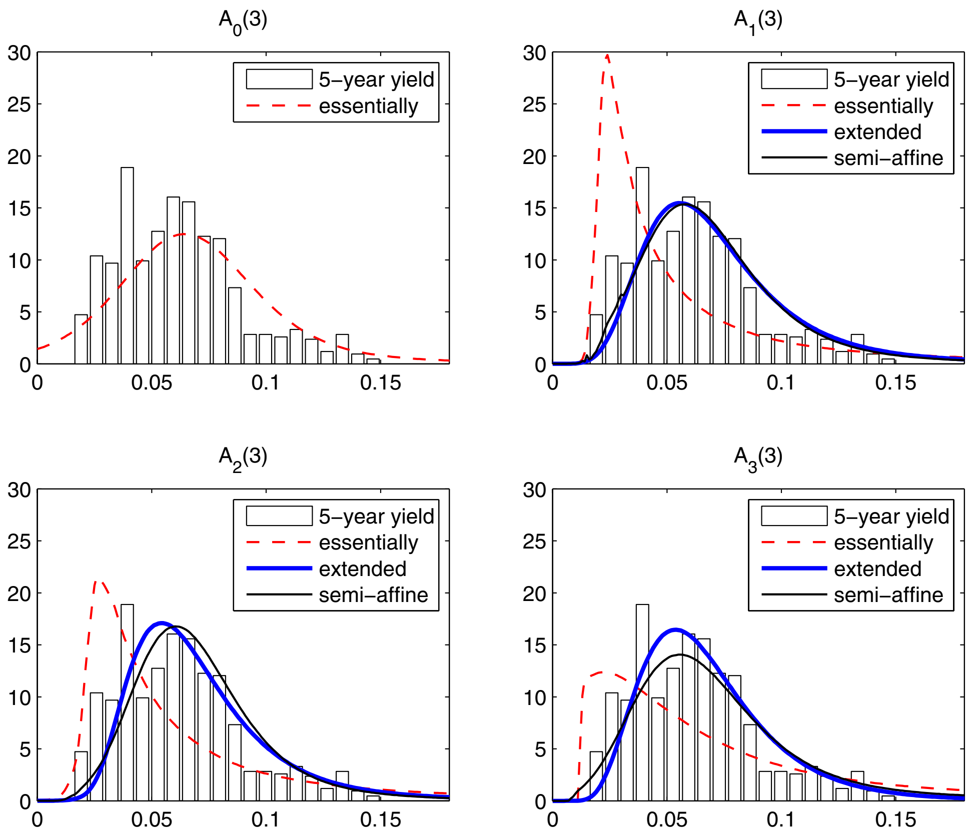

Fig. 2. Distribution of five-year yield. This figure shows the actual distribution of the five-year yield along with model-implied distributions for the essentially, extended, and semi-affine models. Model-implied distributions are based on the estimated values of the cumulative distribution function of the five-year yield for a fine grid of yield values.

under the risk-neutral and historical measure in the drift of each volatility factor. In semi-affine models, the loadings of other volatility factors in the drift of a volatility factor cannot change and therefore the extra risk premium term does not help in capturing the Campbell–Shiller coefficients. However, the extra risk premium term in semi-affine models helps in fitting higher order moments of historical yields. The non-linear term therefore plays a role that is very similar to the extra terms in λ_1 in extended models without requiring the Feller condition to hold.

The Feller condition limits the degree of leptokurtic behavior that yields can exhibit in extended affine models and therefore limits the variety of yield curve shapes. Since the condition is binding extended models are less suited for pricing purposes. To illustrate this phenomenon, Fig. 3 shows the actual

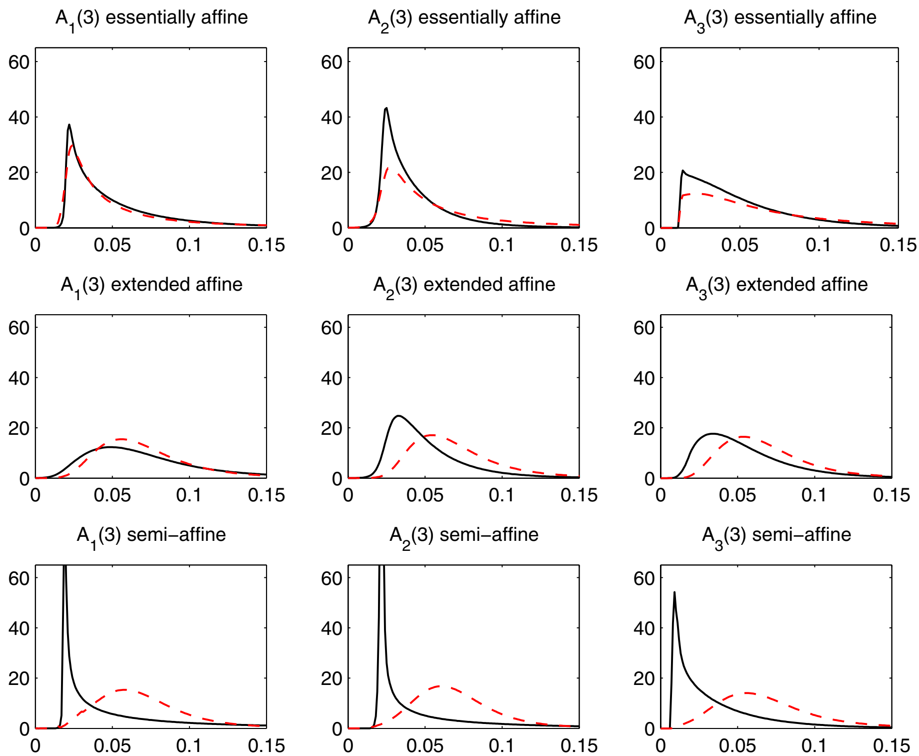

Fig. 3. Distribution of five-year yield under the actual and risk-neutral measure. This figure shows the model-implied distributions under both the historical and risk-neutral measure for the essentially, extended, and semi-affine models. The solid lines show risk-neutral densities while the dashed lines show the actual densities. Model-implied distributions are based on the estimated values of the cumulative distribution function of the five-year yield for a fine grid of yield values.

and risk-neutral distribution of the five-year yield for all models with stochastic volatility.[18]

We see in the top three graphs in the figure that essentially affine models generate actual and risk-neutral distributions that have similar shapes which causes the actual distribution of yields to be too leptokurtic as Fig. 2 documented. Essentially affine models share higher order moments under the historical and risk-neutral measure causing a tension betweeen fitting the shape of the yield curve and the actual distribution of yields. In extended models, we saw that the actual distribution of the five-year yield is fitted well, but the middle three graphs in Fig. 3 show that although the risk-neutral distribution is more

[18] Stationarity is not imposed under the risk-neutral measure, but only the $A_0(3)$ exhibit non-stationarity (in 2.64% of the MCMC draws) and the risk-neutral densities are therefore well defined for all but the $A_0(3)$ model.

leptokurtic than the actual distribution, the leptokurtic behavior is modest and less than in essentially affine models, where the Feller condition is not imposed. The bottom three graphs in the figure shows that in semi-affine models the shapes of the actual and risk-neutral distribution are very different. The models have enough degrees of freedom to fit the actual distribution while allowing the risk-neutral distribution to be strongly skewed and leptokurtic. The shapes of the risk-neutral distributions are similar to those in essentially affine models but the skewness is even more pronounced. This suggests that in essentially affine models the risk-neutral distribution is less leptokurtic than implied solely from the cross section because it shares shape with the actual distribution. Consistent with this, we see that the pricing errors (reflected in σ^2) are smaller for all semi-affine models compared to extended and essentially affine models when comparing models with the same number of volatility factors.

For the affine models with stochastic volatility the cross-sectional fit improves as the skewness and kurtosis of yields increase. This is at odds with the good cross-sectional fit of the Gaussian model — which has zero skewness and kurtosis — and suggests that only once volatility is introduced in the model, high higher order moments improve the cross-sectional fit. In order to investigate what feature of the cross-sectional dimension of the data that leads to high third- and fourth-order moments in stochastic volatility models Table 16 shows the average pricing errors during periods of high volatility (the 10% days with highest EGARCH estimate of monthly volatility of the three-year yield). For all volatility models the one-, three-, and five-year yield is fitted well on average and the pricing errors are insignificantly different from zero. However, affine stochastic volatility models have difficulty matching the curvature of the yield curve precisely when volatility matters the most since they significantly underestimate the two-year yield and overestimate the four-year yield during high volatility. Independent of whether the number of volatility factors is one, two, or three, as higher order moments increase the fit of the two- and four-year yields generally improve.

From the cross section of yields it is not clear what higher moments the risk-neutral distribution of yields have, since the answer depends on the model within which the question is asked. Using options it is possible to provide a model-free estimate of the skewness and kurtosis as done in Bakshi *et al.* (2003) for the equity market. This would provide a model-free test of the affine models considered in this paper, Gaussian models (zero skewness/kurtosis), extended afffine models (only modest skewness/kurtosis), and semi-affine/essentially affine (any skewness/kurtosis). This is an interesting topic for future research.

Table 16. Average pricing errors. This table shows the average pricing errors in basis points for the whole sample period as well as periods with high volatility. Periods of high volatility is determined as the 10% days with highest volatility, where the volatility is an EGARCH(1,1) estimate of monthly volatility of the three-year yield. The stars mark average pricing errors that are significantly different from zero.

Maturity	1	2	3	4	5
			Whole sample period 1952:6 to 2004:12		
$A_0(3)$	−0.05	0.14	0.16	−0.42*	0.15
	(−0.60; 0.50)	(−0.37; 0.61)	(−0.21; 0.52)	(−0.77; −0.06)	(−0.31; 0.64)
$A_1(3)$ ext	−0.14	0.39	−0.04	−0.48*	0.29
	(−0.69; 0.43)	(−0.13; 0.93)	(−0.47; 0.39)	(−0.92; −0.02)	(−0.36; 0.86)
$A_1(3)$ ess	−0.11	0.30	−0.05	−0.41*	0.29
	(−0.68; 0.45)	(−0.13; 0.73)	(−0.40; 0.31)	(−0.82; −0.01)	(−0.29; 0.80)
$A_1(3)$ semi	−0.08	0.19	−0.01	−0.36*	0.22
	(−0.57; 0.44)	(−0.19; 0.57)	(−0.40; 0.38)	(−0.74; −0.02)	(−0.34; 0.71)
$A_2(3)$ ext	−0.18	0.45	−0.08	−0.52*	0.32
	(−0.75; 0.36)	(−0.09; 0.99)	(−0.48; 0.35)	(−1.00; −0.04)	(−0.24; 0.86)
$A_2(3)$ ess	−0.18	0.45	−0.07	−0.51*	0.33
	(−0.75; 0.38)	(−0.06; 0.97)	(−0.46; 0.34)	(−0.94; −0.09)	(−0.22; 0.85)
$A_2(3)$ semi	−0.01	0.06	0.06	−0.34	0.21
	(−0.55; 0.57)	(−0.44; 0.54)	(−0.36; 0.46)	(−0.75; 0.05)	(−0.29; 0.70)
$A_3(3)$ ext	−0.02	0.26	−0.14	−0.48*	0.40
	(−0.61; 0.55)	(−0.30; 0.81)	(−0.55; 0.27)	(−0.89; −0.04)	(−0.12; 0.92)
$A_3(3)$ ess	−0.15	0.41	−0.06	−0.51*	0.30
	(−0.68; 0.41)	(−0.08; 0.90)	(−0.44; 0.34)	(−0.88; −0.12)	(−0.20; 0.83)
$A_3(3)$ semi	−0.09	0.20	0.02	−0.37	0.24
	(−0.63; 0.47)	(−0.33; 0.72)	(−0.40; 0.41)	(−0.80; 0.05)	(−0.31; 0.81)
			Periods with high volatility		
$A_0(3)$	0.12	−0.66	−0.07	1.51*	−1.03
	(−1.58; 1.78)	(−2.00; 0.74)	(−1.09; 0.95)	(0.50; 2.63)	(−2.44; 0.42)
$A_1(3)$ ext	0.90	−2.12*	0.08	2.40*	−1.14
	(−0.84; 2.82)	(−3.48; −0.74)	(−1.15; 1.33)	(1.24; 3.68)	(−2.85; 0.53)
$A_1(3)$ ess	0.77	−2.05*	0.11	2.39*	−1.15
	(−0.87; 2.48)	(−3.32; −0.80)	(−1.01; 1.23)	(1.33; 3.46)	(−2.70; 0.25)
$A_1(3)$ semi	0.47	−1.57*	0.17	2.07*	−1.17
	(−1.22; 2.22)	(−2.75; −0.33)	(−0.99; 1.30)	(0.98; 3.08)	(−2.60; 0.25)
$A_2(3)$ ext	1.09	−2.55*	−0.11	2.55*	−0.94
	(−0.69; 2.74)	(−3.94; −1.16)	(−1.34; 0.98)	(1.35; 3.77)	(−2.55; 0.53)
$A_2(3)$ ess	1.02	−2.24*	−0.19	2.32*	−0.76
	(−0.56; 2.72)	(−3.65; −0.80)	(−1.40; 0.96)	(1.12; 3.49)	(−2.32; 0.95)
$A_2(3)$ semi	0.47	−1.42*	0.00	1.92*	−0.93
	(−1.12; 2.20)	(−2.85; −0.04)	(−1.14; 1.12)	(0.90; 2.97)	(−2.41; 0.51)
$A_3(3)$ ext	0.76	−2.17*	0.35	2.58*	−1.51
	(−1.01; 2.56)	(−3.59; −0.69)	(−0.78; 1.53)	(1.40; 3.78)	(−3.16; 0.01)
$A_3(3)$ ess	0.91	−2.23*	0.04	2.41*	−1.11
	(−0.80; 2.74)	(−3.71; −0.75)	(−1.10; 1.13)	(1.28; 3.56)	(−2.61; 0.39)
$A_3(3)$ semi	0.65	−1.82*	0.13	2.27*	−1.11
	(−1.13; 2.35)	(−3.18; −0.41)	(−1.03; 1.29)	(1.13; 3.42)	(−2.64; 0.38)

5.4. *Zero lower bound*

Given the recent experience of interest rates of zero, there has been a number of papers investigating models that allow for yields close to but not below zero.[19] In the models considered in this paper, only the $A_3(3)$ model with the additional restrictions $\delta_0 \geq 0$ and $\delta_x \geq 0$ guarantee that the short rate is positive, but as noted by Dai and Singleton (2003) a small probability of negative yields is usually accepted in return for a richer correlation structure in affine models with Gaussian variables. Given that recent models with a zero-lower bound typically do not allow for analytical solutions and the computational burden of their estimation is high, it may be worth sacrificing the zero lower bound for tractability reasons if the probability of negative interest rates is small. To test the probability of negative yiels, Table 17 shows the probability that the one- and five-year yields are negative. The table shows that the probability of the one-year yield being negative is less than 1% for all but the $A_0(3)$ model and the probability of a negative five-year yield is practically zero. However, in the Gaussian $A_0(3)$ model the probability of negative one-year yields is 5.98% and the corresponding probability for the five-year yield is 3.91%, probabilities that are many times larger than in the models with stochastic volatility.

Table 17 also shows the probability of seeing extreme values of yields, defined as yields that are higher or lower than what they have been in the estimation period 1952–2004. Due to the skewed and leptokurtic distributions in essentially affine models, the probability of extreme values for the five-year yield ranges from 11.8% to 22.9%. The probabilities are high given that we have not seen these extreme events at any time in the 52 years in the estimation period. In contrast, extended and semi-affine models have probabilities ranging from 2.1% to 6.7% which is a more reasonable range of probabilities.

These results illustrate that the probability of negative yields is high for a purely Gaussian model while it is low for any model with stochastic volatility and the risk premium specification plays a minor role in determining this probability. In contrast, the risk premium specification is important in the probability of observing extreme values of yields and essentially affine models have a high probability of extreme yields while extended and semi-affine models have more reasonable probabilities.

[19] Examples include Kim and Singleton (2012), Priebsch (2013), Wu and Xia (2014), and Christensen and Rudebusch (2015).

Table 17. Probability of negative yields. This table shows for each model the estimated probability of negative yields, smaller yields than experienced in the sample, and higher yields than experienced in the sample. The chosen yields are one- and five-year. The estimated probabilities are based on simulated yields as explained in Appendix B.2.

	$P(Y^1 < 0)$	$P(Y^1 < 0.63\%)$	$P(Y^1 > 15.81\%)$	$P(Y^5 < 0)$	$P(Y^5 < 1.58\%)$	$P(Y^5 > 15.01\%)$
$A_0(3)$ ess	5.98%	7.73%	3.22%	3.91%	7.42%	3.96%
	(0.03%; 38.6%)	(0.06%; 43.8%)	(0.00%; 39.5%)	(0.00%; 31.7%)	(0.05%; 43.6%)	(0.00%; 46.6%)
$A_1(3)$ ess	0.46%	0.97%	8.91%	0.00%	0.80%	11.0%
	(0.00%; 2.00%)	(0.00%; 5.06%)	(0.00%; 66.5%)	(0.00%; 0.00%)	(0.00%; 9.03%)	(0.00%; 68.7%)
$A_2(3)$ ess	0.74%	1.53%	10.7%	0.08%	1.25%	13.2%
	(0.00%; 6.72%)	(0.00%; 12.0%)	(0.00%; 79.6%)	(0.00%; 0.19%)	(0.00%; 7.39%)	(0.00%; 84.1%)
$A_3(3)$ ess	0.00%	0.20%	15.5%	0.00%	4.84%	18.1%
	(0.00%; 0.01%)	(0.00%; 0.57%)	(0.01%; 85.4%)	(0.00%; 0.00%)	(0.02%; 22.7%)	(0.01%; 87.9%)
$A_1(3)$ ext	0.08%	0.34%	3.14%	0.00%	0.11%	4.26%
	(0.00%; 0.39%)	(0.00%; 1.36%)	(0.00%; 30.3%)	(0.00%; 0.00%)	(0.00%; 0.70%)	(0.00%; 36.5%)
$A_2(3)$ ext	0.22%	0.49%	1.84%	0.00%	0.07%	3.16%
	(0.00%; 1.33%)	(0.01%; 2.48%)	(0.00%; 18.5%)	(0.00%; 0.02%)	(0.00%; 0.56%)	(0.00%; 29.4%)
$A_3(3)$ ext	0.02%	0.18%	1.62%	0.00%	0.06%	2.15%
	(0.00%; 0.10%)	(0.01%; 0.88%)	(0.00%; 13.1%)	(0.00%; 0.00%)	(0.00%; 0.46%)	(0.00%; 16.7%)
$A_1(3)$ semi	0.014%	0.52%	3.44%	0.00%	0.20%	4.18%
	(0.00%; 1.10%)	(0.00%; 3.31%)	(0.00%; 41.8%)	(0.00%; 0.00%)	(0.00%; 0.76%)	(0.00%; 46.7%)
$A_2(3)$ semi	0.11%	0.43%	1.53%	0.00%	0.19%	2.47%
	(0.00%; 0.71%)	(0.00%; 2.30%)	(0.00%; 18.7%)	(0.00%; 0.00%)	(0.00%; 1.09%)	(0.00%; 28.3%)
$A_3(3)$ semi	0.02%	0.56%	5.06%	0.00%	0.96%	5.76%
	(0.00%; 0.18%)	(0.00%; 3.83%)	(0.00%; 52.4%)	(0.00%; 0.00%)	(0.00%; 9.62%)	(0.00%; 64.6%)

6. Conclusion

In this paper, I estimate three-factor affine models with different risk premium specifications and examine their ability to match the first four moments of bond yields.

The tension in essentially affine models between fitting time-varying mean and volatility also exists in extended and semi-affine models, although to a lesser extent in extended models. Extended models match historical risk premia better and the improvement increases with the number of stochastic volatility factors, but none of the extended models with stochastic volatility match a purely Gaussian model. However, a purely Gaussian model has no time-varying volatility and I show that it has relatively large probabilities of negative yields. All models capture the broad trends in historical volatility dynamics except in the period of the Fed experiment 1979–1982. The ability to capture time-varying volatility decreases with the number of volatility factors and an affine model with one stochastic volatility factor — essentially, extended, or semi-affine — captures best the historical volatility dynamics.

I document a tension in essentially and extended models in matching both the time series and cross-sectional properties of yields. Essentially affine models fit yields cross sectionally better than extended models but generate historical distributions of yields that are too fat-tailed. The richer risk premium specification in extended affine models allows the historical distribution of yields to be fitted well, but because the Feller condition is imposed extended models cannot generate the variety of yield curve shapes that essentially models can generate and therefore has a worse cross-sectional fit.

Overall, none of the models can fully capture the variation in excess returns and yield volatility. One potential solution to this tension is to increase the number of factors. However, overfitting is a concern when moving beyond three factors as pointed out by Duffee (2010). Another potential solution is to use non-linear models as in Carr *et al.* (2009) and Feldhütter *et al.* (2015). In a non-linear extension of a three-factor Gaussian model, Feldhütter *et al.* (2015) show that the variation in excess returns and yield volatility can be matched.

Acknowledgments

I would like to thank Gregory Duffee, Christian Heyerdahl-Larsen, Philipp Illeditsch, Chris Jones, Scott Joslin, David Lando, Jesper Lund, Mads Stenbo Nielsen, Kenneth Singleton, Anders Bjerre Trolle, and seminar participants at a Berkeley graduate student seminar, the 2007 WHU Campus for Finance Conference, and the 2007 WFA meeting for helpful comments.

Appendix A. Three-Factor Affine Models

In this section, I review the canonical representation of all three-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.

A.1. $A_0(3)$

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

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

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 Eq. (2) has to be non-negative.

A.2. $A_1(3)$

The $A_1(3)$ has the representation

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \left(\begin{bmatrix} K_0^Q(1) \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} K_1^Q(1,1) & 0 & 0 \\ K_1^Q(2,1) & K_1^Q(2,2) & K_1^Q(2,3) \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix}\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}\right) dt$$

$$+\,\mathrm{diag}\left(\begin{bmatrix} \sqrt{X_t^1} \\ \sqrt{1+\beta_2(1)X_t^1} \\ \sqrt{1+\beta_3(1)X_t^1} \end{bmatrix}\right) d\tilde{W}(t).$$

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.[20] For identification the second and third element of δ_x in Eq. (2) has to be non-negative.

The extended affine market price of risk is given as

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

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.[21] 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 the extended affine model nests neither the essential nor the completely affine models.

A.3. $A_2(3)$

The representation of the $A_2(3)$ model is[22]

$$d\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \left(\begin{bmatrix} K_0^Q(1) \\ K_0^Q(2) \\ 0 \end{bmatrix} - \begin{bmatrix} K_1^Q(1,1) & K_1^Q(1,2) & 0 \\ K_1^Q(2,1) & K_1^Q(2,2) & 0 \\ K_1^Q(3,1) & K_1^Q(3,2) & K_1^Q(3,3) \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} \right) dt$$

$$+ \operatorname{diag}\left(\begin{bmatrix} \sqrt{X_t^1} \\ \sqrt{X_t^2} \\ \sqrt{1 + \beta_3(1)X_t^1 + \beta_3(2)X_t^2} \end{bmatrix} \right) d\tilde{W}(t)$$

[20] This parameterization is used in Cheridito *et al.* (2007) and is a consequence of employing the *invariant affine transformation* $T_A X(t) = X(t) + \left(0, \begin{bmatrix} k_{22} & k_{23} \\ k_{32} & k_{33} \end{bmatrix}^{-1} \begin{pmatrix} k_{21} \\ k_{31} \end{pmatrix}\right)'$ 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 $[(K_1^Q)^{-1}K_0^Q]_1 > 0$.

[21] $S_t\Phi_1 + I^-\Phi_2 X_t$ in Eq. (6) can be reparameterized as in Eq. (A.1) 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.

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

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

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

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 $K_0^Q(i) > \frac{1}{2}, i = 1,2$.

A.4. $A_3(3)$

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

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

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

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

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

Appendix B. Implementation Details

B.1. *Model estimation*

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.[23]

The efficiency of the RW-MH algorithm depends crucially on the variance of the proposed 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 eight million draws in the MCMC algorithm. Within each parameter block (K_0^Q, K_1^Q, β, δ, 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 *et al.* (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 100th draw the acceptance ratio of each parameter in a 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).

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

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

[23] According to Gelfand *et al.* (1992) a risk premium element λ can be drawn conditional on the parameter constraint. For example, an element of λ_1 is restricted to $[a;\infty)$ 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.

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

$$\begin{aligned}\mu_t^P &= K_0^Q - K_1^Q X_t + \sqrt{S_t}\Lambda_t \\ &= K_0^Q - K_1^Q X_t + \lambda_0\sqrt{S_t} + \lambda_1 + \lambda_2 X_t,\end{aligned}$$

so in the expression $[\Delta X_t - \mu_{t-1}^P \Delta_t]_i$ all the individual elements λ^{ind} in the vectors λ_0 and λ_1 and matrix λ_2 can be written as $a_t\lambda^{\text{ind}} - b_t$,

$$p(\lambda^{\text{ind}} \mid \Phi_{\setminus \lambda^{\text{ind}}}, X, Y) \propto \exp\left(-\frac{1}{2\Delta_t}\sum_{t=1}^{T}\frac{(a_t\lambda^{\text{ind}} - b_t)^2}{[S_{t-1}]_{ii}}\right).$$

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

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

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 Xs and the sampling of the latent process X can for computational speed be done in two steps. First $X_0, X_2, \ldots$ are sampled and second $X_1, X_3, \ldots$ are sampled. Of the total computing time, solving the ODEs (4)–(5) takes up 70–80% of the computing time.

All random numbers in the estimation are drawn 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.

B.2. *Simulating from models*

The regression coefficients for the Campbell–Shiller and volatility regressions are simulated as follows. 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 50th MCMC draw. This amounts to an average over 400 averages — averaging over a total of 40,000 simulations. To assure that this procedure yields accurate results the following check is performed for the Campbell–Shiller regression coefficients. 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 first procedure on the third decimal and therefore supports that the simulation procedure is accurate.

Unconditional moments and probabilities are calculated using the aforementioned simulation procedure with the exception that for every MCMC draw the estimates are based on one simulation of 30,000 years instead of 100 simulations of 631 months.

Appendix C. Conditional Moments of the CIR Process

According to Cox *et al.* (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

$$\begin{aligned} E(X_{t+\tau}) &= \frac{1}{2c}(f+\lambda) = \theta(1-e^{-k\tau}) + X_t e^{-k\tau}, \\ V(X_{t+\tau}) &= \frac{1}{4c^2}(2f+4\lambda) = \frac{\sigma^2(1-e^{-k\tau})^2}{2k}\theta + \frac{\sigma^2 e^{-k\tau}(1-e^{-k\tau})}{k} X_t, \end{aligned}$$

$$\begin{aligned}\text{skew}(X_{t+\tau}) &= \frac{2^{\frac{3}{2}}(f+3\lambda)}{(f+2\lambda)^{\frac{3}{2}}} = 2^{\frac{3}{2}}\frac{\frac{4k\theta}{\sigma^2}+6cX_te^{-k\tau}}{(\frac{4k\theta}{\sigma^2}+4cX_te^{-k\tau})^{\frac{3}{2}}}\\ &= \frac{\sigma}{\sqrt{k}}\frac{4\theta+\frac{12}{1-e^{-k\tau}}X_te^{-k\tau}}{(2\theta+\frac{4}{1-e^{-k\tau}}X_te^{-k\tau})^{\frac{3}{2}}} = \frac{\sqrt{2}\sigma}{\sqrt{k\theta}}\frac{2\sqrt{2}+\frac{3}{\sqrt{2}}K}{(2+K)^{\frac{3}{2}}},\end{aligned}$$

$$\begin{aligned}\text{exkurt}(X_{t+\tau}) &= \frac{12(f+4\lambda)}{(f+2\lambda)^2} = 12\frac{\frac{4k\theta}{\sigma^2}+8cX_te^{-k\tau}}{(\frac{4k\theta}{\sigma^2}+4cX_te^{-k\tau})^2}\\ &= \frac{12\sigma^2}{k}\frac{4\theta+\frac{16}{1-e^{-k\tau}}X_te^{-k\tau}}{(4\theta+\frac{8}{1-e^{-k\tau}}X_te^{-k\tau})^2} = \frac{3\sigma^2}{k\theta}\frac{4+4K}{(2+K)^2},\end{aligned}$$

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 when X is stationary ($k > 0$). 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 — $\sqrt{\frac{2\sigma^2}{k\theta}}$ and $\frac{3\sigma^2}{k\theta}$.

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, Stanford University.

Almeida, C., J. J. Graveline, and S. Joslin, 2011, Do Interest Rate Options Contain Information About Excess Bond Returns? *Journal of Econometrics* 164, 35–44.

Bakshi, G., N. Kapadia, and D. Madan, 2003, Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *Review of Financial Studies* 16(1), 101–143.

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.

Carr, P., X. Gabaix, and L. Wu, 2009, Linearity-Generating Processes, Unspanned Stochastic Volatility, and Interest-Rate Option Pricing, Working Paper.

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, 2007, Market Price of Risk Specifications for Affine Models: Theory and Evidence, *Journal of Financial Economics* 83, 123–170.

Christensen, J., and G. Rudebusch, 2015, Estimating Shadow-Rate Term Structure Models with Near-Zero Yields, *Journal of Financial Econometrics* 13, 226–259.

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

Cieslak, A. and P. Povala, 2015, Expected Returns in Treasury Bonds, Conditionally Accepted, in *Review of Financial Studies*. http://rfs.oxfordjournals.org/content/early/2015/05/29/rfs.hhv032.

Cochrane, J. H., and M. Piazzesi, 2005, Bond Risk Premia, *American Economic Review* 95, 138–160.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2006, Can Interest Rate Volatility be Extracted From the Yield Curve? An Investigation of Unspanned Stochastic Volatility, Working Paper.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2009, Can Interest Rate Volatility be Extracted From the Cross Section of Bond Yields? *Journal of Financial Economics* 94, 47–66.

Cooper, I., and R. Priestley, 2009, Time-Varying Risk Premiums and the Output Gap, *Review of Financial Studies* 22, 2801–2833.

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

Creal, D., and J. C. Wu, 2015, Estimation of Affine Term Structure Models with Spanned or Unspanned Stochastic Volatility, *Journal of Econometrics* 185, 60–81.

Dai, Q., and K. 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.

Dai, Q., K. J. Singleton, and W. Yang, 2007, Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields, *Review of Financial Studies* 20 (5), 1669–1706.

Duarte, J., 2004, Evaluating an Alternative Risk Preference in Affine Term Structure Models, *Review of Financial Studies* 17(2), 379–404.

Duffee, G., 2010, Sharpe Ratios in Term Structure Models, Working Paper, University of California, Berkeley.

Duffee, G., 2013, Forecasting Interest Rates, in *Handbook of Economic Forecasting*, G. Elliott and A. Timmermann (eds.), Vol. 2A, Chapter 7, Amsterdam: Elsevier, North Holland, pp. 385–426.

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.

Dutta, K. K., and D. F. Babbel, 2002, On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London INter Bank Offer Rates, Working Paper, Wharton.

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.

Feldhütter, P., P. Illeditsch, and C. Heyerdahl-Larsen, 2015, Risk Premia and Volatilities in a Nonlinear Term Structure Model, Working Paper.

Frhwirth-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 Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

Hamilton, J. D. and J. C. Wu, 2012, Identification and Estimation of Gaussian Affine Term Structure Models, *Journal of Econometrics* 168, 315–331.

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.

Jacobs, K., and L. Karoui, 2009, Conditional Volatility in Affine Term-Structure Models: Evidence from Treasury and Swap Markets, *Journal of Financial Economics* 91, 288–318.

Johannes, M., and N. Polson, 2009, MCMC Methods for Financial Econometrics, in *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. Hansen (eds.), Amsterdam, North Holland, Elsevier, Vol. 2, Chapter 1, pp. 1–72.

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

Joslin, S., M. Priebsch, and K. Singleton, 2014, Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, *Journal of Finance* 69(3), 1197–1233.

Joslin, S., K. Singleton, and H. Zhu, 2011, A New Perspective on Gaussian Dynamic Term Structure Models, *Review of Financial Studies* 24(3), 926–970.

Kim, D. H., and K. J. Singleton, 2012, Term Structure Models and the Zero Bound: An Empirical Investigation of Japanese Yields, *Journal of Econometrics* 170(1), 32–49.

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.

Ludvigson, S., and S. Ng, 2009, Macro Factors in Bond Risk Premia, *Review of Financial Studies*, 22(12), 5027–5067.

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

Piazzesi, M., 2010, Affine Term Structure Models, in *Handbook of Financial Econometrics: Tools and Techniques*, Y. Ait-Sahalia and L. Hansen (eds.), Amsterdam, North Holland, Elsevier, Vol. 1, Chapter 12, pp. 691–766.

Priebsch, M., 2013, Computing Arbitrage-Free Yields in Multi-Factor Gaussian Shadow-Rate Term Structure Models, Finance and Economics Discussion Series 2013–63, Board of Governors of the Federal Reserve System (U.S.).

Robert, C. P., and G. Casella, 2004, *Monte Carlo Statistical Methods*, 2nd edn., 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.

Wu, J. C., and F. D. Xia, 2014, Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound, Working Paper.