

Risk Premia and Volatilities in a Nonlinear Term Structure Model

Feldhütter, Peter; Heyerdahl-Larsen, Christian; Illeditsch, Philipp

Document Version

Accepted author manuscript

Published in:

Review of Finance

DOI:

[10.1093/rof/rfw052](https://doi.org/10.1093/rof/rfw052)

Publication date:

2018

License

Unspecified

Citation for published version (APA):

Feldhütter, P., Heyerdahl-Larsen, C., & Illeditsch, P. (2018). Risk Premia and Volatilities in a Nonlinear Term Structure Model. *Review of Finance*, 22(1), 337–380. <https://doi.org/10.1093/rof/rfw052>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jul. 2025



Risk Premia and Volatilities in a Nonlinear Term Structure Model

Peter Feldhütter, Christian Heyerdahl-Larsen, and Philipp Illeditsch

Journal article (Accepted manuscript*)

Please cite this article as:

Feldhütter, P., Heyerdahl-Larsen, C., & Illeditsch, P. (2018). Risk Premia and Volatilities in a Nonlinear Term Structure Model. *Review of Finance*, 22(1), 337-380. DOI: 10.1093/rof/rfw052

This is a pre-copyedited, author-produced version of an article accepted for publication in *Review of Finance* following peer review. The version of record is available online at:

DOI: <https://doi.org/10.1093/rof/rfw052>

* This version of the article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the publisher's final version AKA Version of Record.

Uploaded to [CBS Research Portal](#): February 2019

Risk Premia and Volatilities in a Nonlinear Term Structure Model*

Peter Feldhütter[†]
London Business School

Christian Heyerdahl-Larsen[‡]
London Business School

Philipp Illeditsch[§]
The Wharton School

February 13, 2016

Abstract

We introduce a reduced-form term structure model with closed-form solutions for yields where the short rate and market prices of risk are nonlinear functions of Gaussian state variables. The nonlinear model with three factors matches the time-variation in expected excess returns and yield volatilities of U.S. Treasury bonds from 1961 to 2014. Yields and their variances depend on only three factors, yet the model exhibits features consistent with unspanned risk premia (URP) and unspanned stochastic volatility (USV). The probability of a high volatility scenario increases with the monetary experiment and remains high during the Greenspan area, even though volatilities came back down.

Keywords: Nonlinear term structure models, affine term structure models, expected excess returns, predictability, stochastic volatility, Unspanned Risk Premia (URP), Unspanned Stochastic Volatility (USV), hidden factors.

JEL Classification: D51, E43, E52, G12.

*We would like to thank Kerry Back; Greg Bauer; David Chapman; Mike Chernov; Joao Cocco; Alex David; Greg Duffee; Paul Ehling; Michael Gallmeyer; Francisco Gomes; Rodrigo Guimaraes; Burton Hollifield; Scott Joslin; Christian Julliard; Ralph Koijen; Philippe Mueller; Andreas Pick; Christian Opp; Giuliano De Rossi; David Schröder, Ivan Shaliastovich; Andrea Vedolin; Amir Yaron; seminar participants at London Business School, Wharton School, London School of Economics, University of Melbourne, BI Norwegian Business School, University of Virginia; and participants at the Tripartite Conference held at Wharton, the World Finance Conference in Cyprus, the UBC Summer conference in Vancouver, the Inquire UK conference, and the Eurofidai Finance meeting in Paris for helpful comments and suggestions. This research was funded in part by the Jacobs Levy Equity Management Center for Quantitative Financial Research.

[†]London Business School, Regent's Park, London, NW1 4SA, pfeldhutter@london.edu

[‡]London Business School, Regent's Park, London, NW1 4SA, cheyerdahlarsen@london.edu

[§]The Wharton School, University of Pennsylvania, pille@wharton.upenn.edu

The U.S. Treasury bond market is a large and important financial market. Policy makers, investors, and researchers need models to disentangle market expectations from risk premiums, and estimate expected returns and Sharpe ratios, both across maturity and over time. The most prominent class of models are affine models. However, there are a number of empirical facts documented in the literature that these models struggle with matching *simultaneously*: a) excess returns are time-varying, b) a part of expected excess returns is unspanned by the yield curve, c) yield variances are time varying, and d) a part of yield variances is unspanned by the yield curve.¹ Affine models have been shown to match each of these four findings separately, but not simultaneously and only by increasing the number of factors beyond the standard level, slope, and curvature factors.²

We introduce an arbitrage-free dynamic term structure model where the short rate and market prices of risk are nonlinear functions of Gaussian state variables. We provide closed-form solutions for bond prices and since the factors are Gaussian our nonlinear model is as tractable as a standard Gaussian model. We show that the model can capture all four findings mentioned above *simultaneously* and it does so with only *three* factors driving yields and their variances. The value of having few factors is illustrated by Duffee (2010) who estimates a five-factor Gaussian model to capture time variation in expected returns and finds huge Sharpe ratios due to overfitting.

¹Although the literature is too large to cite in full, examples include Campbell and Shiller (1991) and Cochrane and Piazzesi (2005) on time-varying excess returns, Duffee (2011b) and Joslin, Priebisch, and Singleton (2014) on unspanned expected excess returns, Jacobs and Karoui (2009) and Collin-Dufresne, Goldstein, and Jones (2009) on time-varying volatility, and Collin-Dufresne and Goldstein (2002) and Andersen and Benzoni (2010) on unspanned stochastic volatility.

²Dai and Singleton (2002), and Tang and Xia (2007) find that the only affine three-factor model that can capture time-variation in expected excess returns is the Gaussian model that has no stochastic volatility. Duffee (2011b), Wright (2011) and Joslin, Priebisch, and Singleton (2014) capture unspanned expected excess in four- and five-factor affine models that have no stochastic volatility. Unspanned stochastic volatility is typically modelled by adding additional factors to the standard three factors (Collin-Dufresne, Goldstein, and Jones (2009) and Creal and Wu (2015)). See also Dai and Singleton (2003) and Duffee (2010) and the references therein.

We use a monthly panel of five zero-coupon Treasury bond yields and their realized variances from 1961 to 2014 to estimate the nonlinear model with three factors. To compare the implications of the nonlinear model with those from the standard class of affine models, we also estimate a three-factor affine model with one stochastic volatility factor, the essentially affine $A_1(3)$ model.

We first assess the ability of the nonlinear model to predict excess bond returns in sample and regress realized excess returns on model-implied expected excess return. The average R^2 across bond maturities and holding horizons is 27% for the nonlinear model, 9% for the $A_1(3)$ model, and no more than 15% for any affine model in which expected excess returns are linear functions of yields. Campbell and Shiller (1991) document a positive relation between the slope of the yield curve and expected excess returns, a finding that affine models with stochastic volatility have difficulty matching (see Dai and Singleton (2002)). In simulations, we show that the nonlinear (but not the $A_1(3)$) model can capture this positive relation.

There is empirical evidence that a part of expected excess bond returns is not spanned by linear combinations of yields, a phenomenon we refer to as Unspanned Risk Premia (URP).³ URP arises in our model due to a nonlinear relation between expected excess returns and yields. To quantitatively explore this explanation, we regress expected excess returns implied by the nonlinear model on its Principal Components (PCs) of yields and find that the first three PCs explain 67 – 72% of the variation in expected excess returns. Furthermore, the regression residuals correlate with expected inflation in the data (measured through surveys), not because inflation has any explanatory power in the model but because it happens to correlate with “the amount of nonlinearity.” Duffee (2011b), Wright (2011), and Joslin, Pribsch, and Singleton (2014) use five-factor Gaussian models where one or two factors that

³See Ludvigson and Ng (2009), Cooper and Priestley (2009), Cieslak and Povala (2015), Duffee (2011b), Joslin, Pribsch, and Singleton (2014), Chernov and Mueller (2012), and Bauer and Rudebusch (2015).

are orthogonal to the yield curve explain expected excess returns and are related to expected inflation. We capture the same phenomenon with a nonlinear model that retains a parsimonious three-factor structure to price bonds and yet allows for time variation in volatilities.

The nonlinear and $A_1(3)$ model can capture the persistent time variation in volatilities and the high volatility during the monetary experiment in the early eighties. However, the two models have different implications for the cross-sectional and predictive distribution of yield volatility. In the nonlinear model more than one factor drives the cross-sectional variation in yield volatilities while by construction the $A_1(3)$ model only has one. Moreover, in the nonlinear model the probability of a high volatility scenario increases with the monetary experiment and remains high during the Greenspan area even though volatilities came down significantly. This finding resembles the appearance and persistence of the equity option smile since the crash of 1987. In contrast, the distribution of future volatility in the $A_1(3)$ model is similar before and after the monetary experiment.

There is a large literature suggesting that interest rate volatility risk cannot be hedged by a portfolio consisting solely of bonds; a phenomenon referred to by Collin-Dufresne and Goldstein (2002) as Unspanned Stochastic Volatility (USV). The empirical evidence supporting USV typically comes from a low R^2 when regressing a measure of volatility on interest rates.⁴ To test the ability of the nonlinear model to capture the empirical evidence on USV, we use the methodology of Andersen and Benzoni (2010) and regress the model-implied variance of yields on the PCs of model-implied yields. The first three PCs explain 42 – 44%, which is only slightly higher than in the data where they explain 30 – 35% of the variation in realized yield variance. If we include the fourth and fifth PC, these numbers increase to 55 – 62% and

⁴Papers on this topic include Collin-Dufresne and Goldstein (2002), Heidari and Wu (2003), Fan, Gupta, and Ritchken (2003), Li and Zhao (2006), Carr, Gabaix, and Wu (2009), Andersen and Benzoni (2010), Bikbov and Chernov (2009), Joslin (2014), and Creal and Wu (2014).

40 – 43%, respectively. Hence, our nonlinear model quantitatively captures the R^2 s in USV regressions in the data. In contrast, since there is a linear relation between yield variance and yields in standard affine models, the first three PCs explain already 100% in the $A_1(3)$ model.⁵

The standard procedure in the reduced form term structure literature is to specify the short rate and the market prices of risk as functions of the state variables. Instead, we model the functional form of the stochastic discount factor directly by multiplying the stochastic discount factor from a Gaussian term structure model with the term $1 + \gamma e^{-\beta X}$, where β and γ are parameters and X is the Gaussian state vector. This functional form is a special case of the stochastic discount factor that arises in many equilibrium models in the literature. In such models, the stochastic discount factor can be decomposed into a weighted average of different representative agent models. Importantly, the weights on the different models are time-varying and this is a source of time-varying risk premia and volatility of bond yields.

Our paper is not the first to propose a nonlinear term structure model. Dai, Singleton, and Yang (2007)) estimate a regime-switching model and show that excluding the monetary experiment in the estimation leads their model to pick up minor variations in volatility. In contrast, the nonlinear model can pick up states that did not occur in the sample used to estimate the model. Specifically, we estimate the model using a sample that excludes the monetary experiment and find that it still implies a significant probability of a strong increase in volatility. Furthermore, while the Gaussian model is a special case of both models our nonlinear model only increases the number of parameters from 23 to 27 whereas the regime-switching model in Dai, Singleton, and Yang (2007) has 56 parameters. Quadratic term structure models have

⁵Collin-Dufresne and Goldstein (2002) introduce knife edge parameter restrictions in affine models such that volatility state variable(s) do not affect bond pricing, the so called USV models. The most commonly used USV models—the $A_1(3)$ and $A_1(4)$ USV models—have one factor driving volatility and this factor is independent of yields. These models generate zero R^2 s in USV regressions inconsistent with the empirical evidence.

been proposed by Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2003) among others, but Ahn, Dittmar, and Gallant (2002) find that quadratic term structure models are not able to generate the level of conditional volatility observed for short- and intermediate-term bond yields. Ahn, Dittmar, Gallant, and Gao (2003) propose a class of nonlinear term structure models based on the inverted square-root model of Ahn and Gao (1999), but in contrast to our nonlinear model they do not provide closed-form solutions for bond prices. Dai, Le, and Singleton (2010) develop a class of discrete time models that are affine under the risk neutral measure, but show nonlinear dynamics under the historical measure. They illustrate that the model encompasses many equilibrium models with recursive preferences and habit formation. Carr, Gabaix, and Wu (2009) use the linearity generating framework of Gabaix (2009) to price swaps and interest rate derivatives.

The rest of the paper is organized as follows. Section 1 describes the model. Section 2 estimates the model and Section 3 presents the empirical results. Section 4 concludes.

1 The Model

In this section we present a nonlinear model of the term structure of interest rates. Uncertainty is represented by a d -dimensional Brownian motion $W(t) = (W_1(t), \dots, W_d(t))'$. There is a d -dimensional Gaussian state vector $X(t)$ that follows the dynamics

$$dX(t) = \kappa (\bar{X} - X(t)) dt + \Sigma dW(t), \tag{1}$$

where \bar{X} is d -dimensional and κ and Σ are $d \times d$ -dimensional.

1.1 The Stochastic Discount Factor

We assume that there is no arbitrage and that the strictly positive stochastic discount factor (SDF) is

$$M(t) = M_0(t) \left(1 + \gamma e^{-\beta' X(t)} \right), \quad (2)$$

where γ denotes a nonnegative constant, β a d -dimensional vector, and $M_0(t)$ a strictly positive stochastic process.

Equation (2) is a key departure from standard term structure models (Vasicek (1977), Cox, Ingersoll, and Ross (1985), Duffie and Kan (1996), and Dai and Singleton (2000)). Rather than specifying the short rate and the market price of risk, which in turn pins down the SDF, we specify the functional form of the SDF directly.⁶ This approach is motivated by equilibrium models where the SDF is a function of structural parameters and thus the risk-free rate and market price of risk are interconnected. Moreover, we show in Appendix C that the SDF specified in equation (2) is a special case of the SDF in many popular equilibrium models.

To keep the model comparable to the existing literature on affine term structure models we introduce a base model for which $M_0(t)$ is the SDF. The dynamics of $M_0(t)$ are

$$\frac{dM_0(t)}{M_0(t)} = -r_0(t)dt - \Lambda_0(t)'dW(t), \quad (3)$$

where $r_0(t)$ and $\Lambda_0(t)$ are affine functions of the state vector $X(t)$. Specifically,

$$r_0(t) = \rho_{0,0} + \rho'_{0,X} X(t), \quad (4)$$

$$\Lambda_0(t) = \lambda_{0,0} + \lambda_{0,X} X(t), \quad (5)$$

where $\rho_{0,0}$ is a scalar, $\rho_{0,X}$ and $\lambda_{0,0}$ are d -dimensional vectors, and $\lambda_{0,X}$ is a $d \times d$ -

⁶Constantinides (1992), Rogers (1997), Gabaix (2009), and Carr, Gabaix, and Wu (2009) also specify the functional form of the SDF directly and provide closed form solutions for bond prices.

dimensional matrix. It is well known that bond prices in the base model belong to the class of Gaussian term structure models (Duffee (2002) and Dai and Singleton (2002)) with essentially affine risk premia. If γ or every element of β is zero, then the nonlinear model collapses to the Gaussian base model. We now provide closed form solutions for bond prices in the nonlinear model.

1.2 Closed-Form Bond Prices

Let $P(t, T)$ denote the price at time t of a zero-coupon bond that matures at time T . Specifically,

$$P(t, T) = E_t \left[\frac{M(T)}{M(t)} \right]. \quad (6)$$

We show in the next theorem that the price of a bond is a weighted average of bond prices in artificial economies that belong to the class of essentially affine Gaussian term structure models.

Theorem 1. *The price of a zero-coupon bond that matures at time T is*

$$P(t, T) = s(t)P_0(t, T) + (1 - s(t))P_1(t, T), \quad (7)$$

where

$$s(t) = \frac{1}{1 + \gamma e^{-\beta' X(t)}} \in (0, 1] \quad (8)$$

$$P_n(t, T) = e^{A_n(T-t) + B_n(T-t)' X(t)}. \quad (9)$$

The coefficient $A_n(T - t)$ and the d -dimensional vector $B_n(T - t)$ solve the ordinary

differential equations

$$\frac{dA_n(\tau)}{d\tau} = \frac{1}{2}B_n(\tau)'\Sigma\Sigma'B_n(\tau) + B_n(\tau)'(\kappa\bar{X} - \Sigma\lambda_{n,0}) - \rho_{n,0}, \quad A_n(0) = 0, \quad (10)$$

$$\frac{dB_n(\tau)}{d\tau} = -(\kappa + \Sigma\lambda_{n,X})'B_n(\tau) - \rho_{n,X}, \quad B_n(0) = 0_d, \quad (11)$$

where

$$\rho_{n,0} = \rho_{0,0} + n\beta'\kappa\bar{X} - n\beta'\Sigma\lambda_{0,0} - \frac{1}{2}n^2\beta'\Sigma\Sigma'\beta, \quad (12)$$

$$\rho_{n,X} = \rho_{0,X} - n\kappa'\beta - n\lambda'_{0,X}\Sigma'\beta, \quad (13)$$

$$\lambda_{n,0} = \lambda_{0,0} + n\Sigma'\beta, \quad (14)$$

$$\lambda_{n,X} = \lambda_{0,X}. \quad (15)$$

The proof of this theorem is given in Appendix A (where we provide a proof of a more general class of nonlinear models). To provide some intuition we define $M_1(t) = \gamma e^{-\beta'X(t)}M_0(t)$ and rewrite the bond pricing equation (6) using the fact that $s(t) = M_0(t)/M(t) = 1 - M_1(t)/M(t)$. Specifically,

$$P(t, T) = s(t)E_t\left[\frac{M_0(T)}{M_0(t)}\right] + (1 - s(t))E_t\left[\frac{M_1(T)}{M_1(t)}\right]. \quad (16)$$

Applying Ito's lemma to $M_1(t)$ leads to

$$\frac{dM_1(t)}{M_1(t)} = -r_1(t)dt - \Lambda_1(t)'dW(t), \quad (17)$$

where $r_1(t)$ and $\Lambda_1(t)$ are affine functions of the state vector $X(t)$. Specifically,

$$r_1(t) = \rho_{1,0} + \rho'_{1,X}X(t), \quad (18)$$

$$\Lambda_1(t) = \lambda_{1,0} + \lambda_{1,X}X(t), \quad (19)$$

where $\rho_{1,0}$, $\rho_{1,X}$, $\lambda_{1,0}$, and $\lambda_{1,X}$ are given in equations (12), (13), (14), and (15), respectively. Hence, both expectations in equation (16) are equal to bond prices in artificial economies with discount factors $M_0(t)$ and $M_1(t)$, respectively. These bond prices belong to the class of essentially affine term structure models and hence $P(t, T)$ can be computed in closed form.

1.3 The Short Rate and the Price of Risk

Applying Ito's lemma to equation (2) leads to the dynamics of the SDF:

$$\frac{dM(t)}{M(t)} = -r(t) dt - \Lambda(t)' dW(t), \quad (20)$$

where both the short rate $r(t)$ and the market price of risk $\Lambda(t)$ are nonlinear functions of the state vector $X(t)$ given in equations (21) and (22), respectively. The short rate is given by

$$r(t) = s(t)r_0(t) + (1 - s(t))r_1(t). \quad (21)$$

Our model allows the short rate to be nonlinear in the state variables without losing the tractability of closed form solutions of bond prices and a Gaussian state space.⁷

The d -dimensional market price of risk is given by

$$\Lambda(t) = s(t)\Lambda_0(t) + (1 - s(t))\Lambda_1(t). \quad (22)$$

Equation (22) shows that even if the market prices of risk in the base model are constant, the market prices of risks in the general model are stochastic due to varia-

⁷Chan, Karolyi, Longstaff, and Sanders (1992), Ait-Sahalia (1996a), Ait-Sahalia (1996b), Stanton (1997), Pritsker (1998), Chapman and Pearson (2000), Ang and Bekaert (2002), and Jones (2003) study the nonlinearity of the short rate. Jermann (2013) and Richard (2013) study nonlinear term structure models, but they do not obtain closed form solutions for bond prices.

tions in the weight $s(t)$. When $s(t)$ approaches zero or one, then $\Lambda(t)$ approaches the market price of risk of an essentially affine Gaussian model.

1.4 Expected Return and Volatility

We know that the bond price is a weighted average of exponential affine bond prices (see equation (7)). Hence, variations of instantaneous bond returns are due to variations in the two artificial bond prices $P_0(t, T)$ and $P_1(t, T)$ and due to variations in the weight $s(t)$. Specifically, the dynamics of the bond price $P(t, T)$ are

$$\frac{dP(t, T)}{P(t, T)} = (r(t) + e(t, T)) dt + \sigma(t, T)' dW(t), \quad (23)$$

where $e(t, T)$ denotes the instantaneous expected excess return and $\sigma(t, T)$ denotes the local volatility vector of a zero-coupon bond that matures at time T .

The local volatility vector of the bond is given by

$$\sigma(t, T) = \omega(t, T)\sigma_0(T - t) + (1 - \omega(t, T))\sigma_1(T - t) + (s(t) - \omega(t, T))\beta, \quad (24)$$

where $\sigma_i(T - t) = \Sigma' B_i(T - t)$ denotes the local bond volatility vector in the Gaussian model with SDF $M_i(t)$ and $\omega(t, T)$ denotes the contribution of $P_0(t, T)$ to the bond price $P(t, T)$. Specifically,

$$\omega(t, T) = \frac{P_0(t, T)s(t)}{P(t, T)} \in (0, 1]. \quad (25)$$

When $s(t)$ approaches zero or one, then $\sigma(t, T)$ approaches the deterministic local volatility of a Gaussian model. However, in contrast to the short rate and the market price of risk, the local volatility can move outside the range of the two local Gaussian volatilities, $\sigma_0(T - t)$ and $\sigma_1(T - t)$, because of the last term in equation (24).

Intuitively, there are two distinct contributions to volatility in equation (24). The direct term, defined as

$$\sigma_{vol}(t, T) = \omega(t, T)\sigma_0(T - t) + (1 - \omega(t, T))\sigma_1(T - t), \quad (26)$$

arises because the two artificial Gaussian models have constant but different yield volatilities. The indirect term, defined as

$$\sigma_{lev}(t, T) = (s(t) - \omega(t, T))\beta \quad (27)$$

is due to the Gaussian models having different yield levels. Two special cases illustrate the distinct contributions to volatility. If $P_0(t, T) = P_1(t, T) = P(t, T)$, then $\sigma_{lev}(t, T) = 0$ and the local volatility vector reduces to $\sigma(t, T) = s(t)\sigma_0(T - t) + (1 - s(t))\sigma_1(T - t)$. On the other hand, if $\sigma_0(T - t) = \sigma_1(T - t)$, the first term is constant, but there is still stochastic volatility due to the second term which becomes more important the bigger the difference between the two artificial bond prices $P_1(t, T)$ and $P_0(t, T)$.⁸

The instantaneous expected excess return and volatility of the bond are

$$e(t, T) = \Lambda(t)'\sigma(t, T) \quad (28)$$

$$v(t, T) = \sqrt{\sigma(t, T)'\sigma(t, T)}. \quad (29)$$

Equations (20)-(29) show that the nonlinear term structure model differs from the essentially affine Gaussian base model in two important aspects. First, the volatilities of bond returns and yields are time-varying and hence expected excess returns are moving with both the price and the quantity of risk.⁹ Second, the short rate $r(t)$, the

⁸If $\lambda_{0,X}$ and κ are zero, then $\sigma_0(T - t) = \sigma_1(T - t)$.

⁹The instantaneous volatility of the bond yield is $\frac{1}{T}v(t, t + \tau)$.

instantaneous volatility $v(t, T)$, and the instantaneous expected excess return $e(t, T)$ are nonlinear functions of $X(t)$.

2 Estimation

In this section, we estimate the nonlinear model described in Section 1 and compare it to a standard essentially affine $A_1(3)$ model. Both models have three factors and the number of parameters is 23 in the affine model and 26 in the nonlinear model. The $A_1(3)$ model is well known and thus we defer details to Appendix B.

2.1 Data

We treat each period as a month and estimate the models using a monthly panel of five zero-coupon Treasury bond yields and their realized variances.¹⁰ We use daily (continuously compounded) 1-, 2-, 3-, 4-, and 5-year zero-coupon yields extracted from U.S. Treasury security prices by the method of Gurkaynak, Sack, and Wright (2007). The data is available from the Federal Reserve Board's webpage and covers the period 1961:07 to 2014:04. For each bond maturity, we average daily observations within a month to get a time series of monthly yields. We use realized yield variance to measure yield variance. Let y_t^τ and rv_t^τ denote the yield and realized yield variance of a τ -year bond in month t based on daily observations within that month. Specifically,

$$y_t^\tau = \frac{1}{N_t} \sum_{i=1}^{N_t} y_{d,t}^\tau(i), \quad (30)$$

$$\text{rv}_t^\tau = 12 \sum_{i=1}^{N_t} \left(y_{d,t}^\tau(i) - y_{d,t}^\tau(i-1) \right)^2, \quad (31)$$

¹⁰Cieslak and Povala (2014) also include realized variance in the estimation of term structure models.

where $y_{d,t}^\tau(i)$ denotes the yield at day i within month t , N_t denotes the number of trading days within month t , and $y_{d,t}^\tau(0)$ denotes the last observation in month $t - 1$. The realized variance converges to the quadratic variation as N approaches infinity, see Andersen, Bollerslev, and Diebold (2010) and the references therein for a detailed discussion.

To check the accuracy of realized variance based on daily data, we compare realized volatility with option-implied volatility (to be consistent with the options literature we look at implied volatility instead of implied variance). We obtain implied price volatility of one month at-the-money options on five-year Treasury futures from Datastream and convert it to yield volatility.¹¹ We then calculate monthly volatility by averaging over daily volatilities. Figure 1 shows that realized volatility tracks option-implied volatility closely (the correlation is 87%), and thus we conclude that realized variance is a useful measure for yield variance.

2.2 Estimation Methodology

We use the Unscented Kalman Filter (UKF) to estimate the nonlinear model and the approximate Kalman filter to estimate the $A_1(3)$ model. Christoffersen, Dorion, Jacobs, and Karoui (2014) show that the UKF works well in estimating term structure models when highly nonlinear instruments are observed. We briefly discuss the setup but refer to Christoffersen, Dorion, Jacobs, and Karoui (2014) and Carr and Wu (2009) for a detailed description of this nonlinear filter.

We stack the five yields in month t in the vector Y_t , the corresponding five realized yield variances in the vector RV_t , and set up the model in state-space form. The

¹¹We calculate yield volatility by dividing price volatility with the bond duration. We calculate bond duration in two steps. We first find the coupon that makes the present value of a five year bond's cash flow equal to the at-the-money price of the underlying bond the option is written on (available from Datastream). We then calculate the modified duration of this bond.

measurement equation is

$$\begin{pmatrix} Y_t \\ \text{RV}_t \end{pmatrix} = \begin{pmatrix} f(X_t) \\ g(X_t) \end{pmatrix} + \begin{pmatrix} \sigma_y I_5 & 0 \\ 0 & \sigma_{rv} I_5 \end{pmatrix} \epsilon_t, \quad \epsilon_t \sim N(0, I_{10}), \quad (32)$$

where $f(\cdot)$ is the function determining the relation between the latent variables and yields, $g(\cdot)$ is the function determining the relation between the latent variables and the variance of yields, and the positive parameters σ_{rv} and σ_y are the pricing errors for yields and their variances.¹² Specifically, $f = (f_1, \dots, f_5)'$ and $g = (g_1, \dots, g_5)'$ where

$$f_\tau(X_t) = -\frac{1}{\tau} \ln(P(X_t, t + \tau)) \quad (33)$$

$$g_\tau(X_t) = \frac{1}{\tau^2} v^2(X_t, t + \tau) \quad (34)$$

with $P(X_t, t + \tau)$ and $v(X_t, t + \tau)$ given in Equation (7) and (29), respectively.

In the nonlinear model the state space is Gaussian and thus the transition equation for the latent variables is

$$X_{t+1} = C + DX_t + \eta_{t+1}, \quad \eta_t \sim N(0, Q), \quad (35)$$

where C is a vector and D is a matrix that enters the one-month ahead expectation of X_t , i.e., $E_t(X_{t+1}) = C + DX_t$. The covariance matrix of X_{t+1} given X_t is constant and equal to Q .

In the $A_1(3)$ model we use the Gaussian transition equation in (35) as an approximation because the dynamics of X are non-Gaussian. This is a standard approach in the literature (Feldhütter and Lando (2008)). The bond price $P(X_t, t + \tau)$ and

¹²We choose to keep the estimation as parsimonious as possible by letting the σ_{rv} be the same for all realized variances. An alternative is to use the theoretical result in Barndorff-Nielsen and Shephard (2002) that the variance of the measurement noise is approximately two times the square of the spot variance and allow for different measurement errors across bond maturity.

volatility $v(X_t, t + \tau)$ in equation (33) and (34) of the $A_1(3)$ model are given in equation (61) and (62) in Appendix B. We can use the approximate Kalman filter because both yields and variances are affine in X in the $A_1(3)$ model.

We use the normalization proposed in Dai and Singleton (2000) to guarantee that the parameters are well identified if $s(X_t)$ is close to zero or one, or if γ and all elements of β are close to zero. In the nonlinear model, we assume in Equation (1) that the mean reversion matrix, κ , is lower triangular, the mean of the state variables, \bar{X} , is the zero vector, and that the local volatility, Σ , is the identity matrix. The normalizations in the $A_1(3)$ model are given in Appendix B.

2.3 Estimation Results

Estimated parameters are reported in Table 1. The volatility of the pricing errors σ_y and σ_{rv} show that the nonlinear model matches yields slightly better and realized variances slightly worse, but overall the fit of both models is of similar magnitude. Figure 2 shows that both models fit yields well. Figure 3 shows that there are larger fitting errors for realized variances than for yields, which is not surprising given realized variance is a noisy estimate of the true underlying variance.

The bond price in the nonlinear model is a weighted average of two Gaussian bond prices (see Theorem 1). Figure 4 shows the weight $s(X_t)$ on the Gaussian base model. If the stochastic weight approaches zero or one, then the bond price approaches the bond price in a Gaussian model where yields are affine functions of the state variables and yield variances are constant. The stochastic weight is distinctly different from one and varies substantially over the sample period, that is, the mean and volatility of $s(X_t)$ are 79.98% and 21.35%, respectively. Moreover, there are both high-frequency and low-frequency movements in $s(X_t)$. The high-frequency movements push $s(X_t)$ away from one during recessions; we see spikes during the 1970, 1973-1975, 1980, 2001,

and 2007-2009 recessions. The low-frequency movement start in the early eighties where the weight moves significantly below one and slowly returns over the next 30 years.

To quantify the impact of nonlinearities in our model, we regress yields and their variances on the three state variables. By construction the R^2 of these regressions in the $A_1(3)$ model is 100%. In the nonlinear model, the R^2 s when regressing the one to five-year yields on the three state variables are 89.40%, 89.64%, 90.12%, 90.66%, and 91.14%, respectively, showing a considerable amount of nonlinearity. Nonlinearity shows up even stronger in the relation between yield variances and the three factors. Specifically, the R^2 s when regressing the one to five-year yield variances on the three state variables are 29.52%, 27.99%, 28.18%, 29.52%, and 31.67%, respectively. For comparison, regressing the stochastic weight $s(X_t)$ on all three state variables leads to an R^2 of 80.88%. Overall, these initial results suggest an important role for nonlinearity and we explore this in detail in the next section.

3 Empirical Results

In this section we focus on the empirical properties of the nonlinear model and compare it to a standard affine model with stochastic volatility, the $A_1(3)$ model.

3.1 Expected Excess Returns

Expected excess returns of U.S. Treasury bonds vary over time as documented in among others Fama and Bliss (1987) and Campbell and Shiller (1991) (CS). CS document this by regressing future yield changes on the scaled slope of the yield curve. The slope regression coefficient is one if excess holding period returns are constant, but CS find negative regression coefficients implying that a steep slope predicts high

future excess bond returns. Panel A in Table 2 shows that the nonlinear model captures the negative CS regression coefficients in population, while the $A_1(3)$ model does not. This is consistent with evidence in Dai and Singleton (2002), Tang and Xia (2007), and Feldhütter (2008) that in the class of three-factor affine models the $A_1(3)$ model cannot generate negative CS regression coefficients.¹³

Figure 5 shows that one-year expected excess returns in the nonlinear model are positive since the mid-80s while they are alternating between positive and negative in the $A_1(3)$ model. Although not shown in the figure, this is also the case at longer holding horizons. Realized returns are predominantly positive since the mid-80s, which may suggest that the nonlinear model fits expected excess returns better than the $A_1(3)$ model. We run a regression of realized excess returns on expected excess returns to test this formally.¹⁴ The results are reported in Panel B of Table 2. If the model captures expected excess returns well, then the slope coefficient should be one, the constant zero, and the R^2 high. The slope coefficients are lower but generally close to one in the nonlinear model with an average slope coefficient of 0.85. In the $A_1(3)$ model the slope coefficients are basically one at the one-year horizon but become too low at longer horizons and the average coefficient is 0.69. Similarly, the average constant α is closer to zero in the nonlinear model (-0.0046) than in the $A_1(3)$ model (0.0089). The average R^2 across bond maturity and holding horizon is 27% in the nonlinear model while it is only 9% in the $A_1(3)$ model.

To measure how well the nonlinear model predicts excess returns we compare the mean squared error of the predictor to the unconditional variance of excess returns. Specifically, we define the statistic “fraction of variance explained” that measures the

¹³The mentioned studies show that only the Gaussian model can match the CS regression coefficients, while the $A_1(3)$ comes closest to matching the coefficients among the models with stochastic volatility, $A_n(3)$, $n > 0$. We do not compare with a Gaussian model because it does not allow for stochastic volatility and therefore it fails in capturing the time-variation in yield volatility.

¹⁴Moments of yields and returns are easily calculated using Gauss-Hermite quadrature, see Appendix D for details. In the rest of the paper we use Gauss-Hermite quadrature when we do not have closed-form solutions for expectations or variances.

explanatory power of the model implied expected excess return as follows¹⁵

$$\text{FVE} = 1 - \frac{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - E_t[rx_{t,t+n}^\tau])^2}{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - \frac{1}{T} \sum_{t=1}^T rx_{t,t+n}^\tau)^2}, \quad (36)$$

where $rx_{t,t+n}^\tau$ is the n -year log return on a bond with maturity τ in excess of the n -year yield and $E_t[rx_{t,t+n}^\tau]$ is the corresponding model implied expected excess return. If the predictor is unbiased, then the R^2 from the regression of realized on expected excess returns is equal to the FVE and otherwise it is an upper bound. Panel B of Table 2 shows that the FVEs in the nonlinear model are higher than in the $A_1(3)$ model, and in contrast to the nonlinear model, the performance of the $A_1(3)$ model deteriorates as we increase the holding horizon.

To compare the nonlinear model to affine models more generally we regress future excess returns on the five yields. The R^2 s from this regression corresponds to the explanatory power of the Cochrane and Piazzesi (2005) factor and is an upper bound for the FVE of any affine model for which expected excess returns are spanned by yields. Panel B of Table 2 shows that the FVEs of the nonlinear model are equal to or higher than the explanatory power of the Cochrane-Piazzesi factor.¹⁶ This implies that no affine model without hidden risk premium factors (see discussion below) can explain more of the variation in realized excess returns than the nonlinear model.

¹⁵Almeida, Graveline, and Joslin (2011) refer to this measure as a modified R^2 .

¹⁶The average R^2 from regressing excess returns onto yields for a one-year holding horizon is 17% which is lower than the 37% reported in Cochrane and Piazzesi (2005). There are two reasons for this. First, the data sets are different. If we use the Fama-Bliss data, then the average R^2 increases to 25%. Second, Cochrane and Piazzesi (2005) use the period 1964-2003 and R^2 's are lower outside this sample period as documented in Duffee (2012).

3.1.1 Unspanned Risk Premia

There is a lot of empirical evidence that shows that a part of excess bond returns is explained by macro factors not spanned by linear combinations of yields.¹⁷ For example, Bauer and Rudebusch (2015) find that the R^2 when regressing realized excess returns on the first three PC of yields along with expected inflation is 85% higher than when regressing on just the first three PCs.¹⁸ We refer to this empirical finding as Unspanned Risk Premia or URP.

To quantitatively capture URP in a term structure model Duffee (2011b), Joslin, Priebisch, and Singleton (2014), and Chernov and Mueller (2012) use five-factor Gaussian models. The reason for using five factors is that three factors are needed to explain the cross section of bond yields and then one or two factors orthogonal to the yield curve explain expected excess returns. An alternative explanation for the spanning puzzle that has not been explored in the literature is that there is a nonlinear relation between yields and expected excess returns. We therefore ask the question: are nonlinearities empirically important for understanding the spanning puzzle?

To answer the question, we start by regressing model-implied one-year expected excess return on the first five PCs of model-implied yields in the sample period. The R^2 s of these regressions are reported in Table 3. Panel C shows that by construction the first three PCs explain all the variation in expected excess returns in the $A_1(3)$ model (since expected excess returns are linear in the yields). Panel B shows that the first three PCs explain on average 69.4% of the variation of expected excess returns in the nonlinear model in the sample period. That is, almost one third of the variation of expected excess returns is due to a nonlinear relation between expected excess

¹⁷See Ludvigson and Ng (2009), Cooper and Priestley (2009), Cieslak and Povala (2015), Duffee (2011b), Joslin, Priebisch, and Singleton (2014), and Chernov and Mueller (2012). Bauer and Rudebusch (2015) argue that this evidence can be explained by measurement error.

¹⁸The R^2 is 0.36 in the former and 0.195 in the latter, see Bauer and Rudebusch (2015)'s Table 3. Joslin, Priebisch, and Singleton (2014) present similar evidence.

returns and yields.

Empirically, realized excess returns are invariably used in lieu of expected excess returns as dependent variable. To see how this noise affects the importance of nonlinearities, we simulate 1,000,000 months and regress realized excess returns on the five PCs. Panel D shows the results for the nonlinear model, while Panel E shows the results for the $A_1(3)$ model. We see that the R^2 in the nonlinear model are largely in line with the actual R^2 in Panel A while this is not the case for the $A_1(3)$ model. The final column shows the R^2 s when regressing realized excess returns on expected excess returns. The average R^2 is 81% higher than when regressing on the first three PCs in the nonlinear model (26.2% vs 14.5%). This implies that if there is a macro variable that perfectly tracks expected excess returns, average R^2 s when regressing realized excess returns on the first three PCs and this macro factor would be 81% higher than when regressing on just the first three PCs; similar to the incremental R^2 documented in Bauer and Rudebusch (2015). Of course, this is not because this macro factor contains any information not in the yield curve.

Is it plausible that macro factors (partially) pick up nonlinearities? To address this question, we take the residuals from regressing expected excess returns on PCs in the nonlinear model (Panel B in Table 3) and regress them on expected inflation.¹⁹ Table 4 shows the results. Expected inflation explains about 11% of the variation and is statistically significant at the 5% level when using residuals based on the first three PCs. The R^2 s increase to slightly less than 20% when adding the fourth PC. Even when including all five PCs expected inflation is highly significant. Thus, although all information about expected excess returns is contained in the yield curve, expected inflation appears to contain information about them when running linear regressions.

¹⁹Expected inflation is measured as the mean forecasts of price growth in the Michigan Survey of Consumers (MSC). MSC is a survey conducted on monthly frequencies covering a large cross-section of consumers and Ang, Bekaert, and Wei (2007) show that it is a good unbiased predictor of inflation.

Overall, our nonlinear model highlights an alternative channel that helps explain the spanning puzzle: expected excess returns are nonlinearly related to yields and therefore a part of expected excess returns appears to be “hidden” from a linear combination of yields and this part can be picked up by macro factors. This is achieved in a parsimonious three-factor model rather than a five-factor model as is common in the literature.

3.2 Stochastic Volatilities

Table 6 shows that there is more than one factor in realized yield variances in our data: the first PC of yield variances explain 94,5% of the variation while the first two PCs explain 99.2%. The $A_1(3)$ model has by definition only one factor explaining volatilities and therefore the first PC explain all the variation in model-implied realized variances.²⁰ In the nonlinear model, the first PC explains 97.5% of the variation in model-implied variances and the first two PCs explain 99.9%. Hence, yield variances in the nonlinear model exhibit a linear multi-factor structure as in the data.

The nonlinear and $A_1(3)$ model also have significantly different distributions of future yield volatility. Figure 6 shows the one-year ahead conditional distribution of the instantaneous yield volatility for the bond with three years to maturity (the distributions for bonds with other maturities are similar).²¹ The volatility is a linear function of only one factor in the $A_1(3)$ model and the distribution of future volatility is fairly symmetric and does not change much over time. In the nonlinear model volatility is a nonlinear function of three factors and the volatility distribution takes on a variety of shapes that persist over time.

The 97.5 quantiles of the one-year ahead volatility distribution in the nonlinear

²⁰Even though realized variances are noisy measures of integrated variances, average yields nevertheless span realized variances, see Andersen and Benzoni (2010).

²¹The instantaneous yield volatility is $\frac{1}{T}v^{(\tau)}(t)$ with $v^{(\tau)}(t)$ given in equation (29).

model shows that the market did not anticipate the possibility of very volatile yields before the monetary experiment in the early 80s, apart from brief periods around the 1970s recessions. However, there is a significant probability of a high yield volatility scenario since the 80s, despite the fact that volatilities have come down to levels similar to those in the 60s and 70s. It is only in the calm 2005-2006 period a high-volatility scenario was unlikely. This finding suggest that there is information about the risk of a high volatility regime in Treasury bond data which is similar to the appearance of the smile in equity options since the stock market crash of 87. To test whether the nonlinear model can capture this information from Treasury bond data that exclude the monetary experiment, we re-estimate the parameters using yield and realized yield variance data for the period August 1987 to April 2014 (Alan Greenspan became chairman of the Fed on August 11, 1987). Table 1 shows the parameters of the nonlinear model estimated using yields and realized yield variances for the period August 1987 to April 2014. The parameters β and γ that capture the nonlinearities of the model are similar to the estimates using data that includes the early 80s. Figure 7 shows the 97.5 quantiles of the one-year ahead distribution of yield volatility for sample periods with and without the early 80s. There is a fat right-tail in the volatility distribution in both cases and hence the nonlinear model captures the risk of strong increase in volatility, even when such an event is not in the sample used to estimate the model.

The regime-switching models of Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), and Bansal, Tauchen, and Zhou (2004) capture time variation in the probabilities of high volatility regimes by adding a state variable that picks up the regime. However, if a high-volatility regime is not in the sample used to estimate the model, then the regimes in the model will pick up minor variations in volatility (see the discussion in Dai, Singleton, and Yang (2007)). Everything works through nonlinearities in our model and therefore the probability of a high-volatility regime

can be pinned down in a sample that does not include such an episode.

3.2.1 Unspanned Stochastic Volatility

There is a large literature suggesting that interest rate volatility risk cannot be hedged by a portfolio consisting solely of bonds; a phenomenon referred to by Collin-Dufresne and Goldstein (2002) as Unspanned Stochastic Volatility (USV). The empirical evidence supporting USV typically comes from a low R^2 when regressing measures of volatility on interest rates. For instance, Collin-Dufresne and Goldstein (2002) regress straddle returns on changes in swap rates and document R^2 s as low as 10%. Similarly, Andersen and Benzoni (2010) (AB) regress yield variances - measured using high frequency data - on the first six PCs of yields and find low R^2 s. Inconsistent with this evidence, standard affine models produce high R^2 's in USV regressions because there is a linear relation between yield variances and yields in the model.

The nonlinear model provides an alternative explanation for low R^2 s in USV regression because the relation between yield variances and yields is nonlinear. However, it is an empirical question if nonlinearities in the model are strong enough to produce R^2 s similar to those found in the data. To answer this question, we follow AB and regress realized yield variance on principal components of yields. The R^2 s of these regressions for the data are reported in Panel A of Table 5. The average R^2 when regressing realized variance on the first three PCs is 32.4%, confirming that the PCs of yields only explain a fraction of the variation in variance.²² Panel B shows that the average R^2 is 42.5% when we regress monthly model-implied instantaneous yield variance on the first three PCs of monthly model-implied yields which is not substantially higher than in the data. In contrast, Panel D shows that in the $A_1(3)$ model the R^2 is 100% once the first three PCs are included in the regression of in-

²²The R^2 are higher than those found in AB because the sample period includes the monetary experiment, see Jacobs and Karoui (2009) for a discussion of the explanatory power in USV regressions for different time periods.

stantaneous variance on yield PCs. Hence, the presence of nonlinearities give rise to low R^2 's in AB's USV regression.

To understand why a significant part of variance is (linearly) unspanned by yields we recall that equation (24) shows that the local volatility consists of two components, σ_{lev} and σ_{vol} , and thus the instantaneous yield variance is

$$\sigma(t, T)' \sigma(t, T) = \sigma_{vol}(t, T)' \sigma_{vol}(t, T) + \sigma_{lev}(t, T)' \sigma_{lev}(t, T) + 2\sigma_{vol}(t, T)' \sigma_{lev}(t, T) \quad (37)$$

While the average R^2 across maturities when regressing the yield variance on the first five PCs of model-implied yields is only 59.2%, the average R^2 when regressing each component in (37) on the five PCs of yields is 94.4%, 88.2%, and 94.9%, respectively. Hence, each component is close to being linearly spanned, but they partially offset each other.²³ When $P_1(t, T) = P_2(t, T)$ the second and third term in (37) vanish and volatility is largely spanned. Hence, the fraction of volatility that is unspanned varies significantly over time consistent with findings in Jacobs and Karoui (2009).

Bikbov and Chernov (2009) discuss how measurement error due to microstructure effects such as the bid-ask spread in option and bond prices affects the explanatory power of USV regressions. Collin-Dufresne and Goldstein (2002) argue that measurement error cannot be the reason for low R^2 's in USV regressions because there is a strong factor structure in the regression residuals across bond maturities. Panel F of Table 5 confirms the factor structure in the data because the first PC explains 91.8% of the residual variation. The first PC explains 98% of the residual variation in the regression implied by the nonlinear model. Hence, our nonlinear model can capture the low explanatory power and the strong residual factor structure of the USV regressions that is observed in the data.

²³In particular, as $s(t)$ moves towards the high volatility model, the yield difference between the two models tends to decrease. That is, as the first part in (37) increases, the second part in the same equation tends to decrease.

The USV regressions in Table 5's Panel A are subject to a measurement error not discussed in Bikbov and Chernov (2009) due to the use of realized variance instead of option-implied variance. To assess the importance of this measurement error we consider the nonlinear and $A_1(3)$ model and simulate 1,000,000 months of daily data (with 21 days in each month), compute monthly realized variance and monthly average yield, and regress realized variance on the five PCs of yields. Panel C shows the results for the nonlinear model. The results are very similar to those when using instantaneous variance in the sample period – the average R^2 is 39.8% for the first three PCs – so our results are robust to taking into account measurement error in realized variance. Panel E shows that the average R^2 is 45.8% in the $A_1(3)$ model when regressing realized variance on the first three PCs of yields in the simulated sample, which brings model R^2 s much closer to data R^2 s. However, the R^2 s when using only one or two PCs in the model are zero which is strongly at odds with the data.²⁴

Collin-Dufresne and Goldstein (2002) introduce knife edge parameter restrictions in affine models such that volatility state variable(s) do not affect bond yields, the so-called USV models. The most commonly used USV models – the $A_1(3)$ and $A_1(4)$ – have one factor driving volatility and this factor does not affect yields. These models generate zero R^2 s in the above USV regression in population, inconsistent with the empirical evidence. In contrast, the nonlinear model retains a parsimonious three-factor structure and yet can generate R^2 s in USV regressions which are broadly in line with those in the data.

²⁴Since measurement errors when using realized variance in the $A_1(3)$ model result in a drop in R^2 s from 100% to 45.8%, an interesting question is if the population R^2 s in the nonlinear model in Panel C would be substantially higher if instantaneous variance is used instead of realized variance. The answer is no. If instantaneous model-implied variance is used the average R^2 is 48.4% instead of 43.6% in Panel C.

3.3 Cross-sectional fit of three-factor models

The nonlinear bond pricing model allows us to capture the observed time variation in the mean and volatility of excess bond returns. However, Balduzzi and Chiang (2012) show that in the cross-section there is an almost linear relation between yields of different maturities. To check whether the nonlinear model captures the cross-sectional linearity we follow Duffee (2011a) and determine the principal components of zero-coupon bond yield changes with maturities ranging from one to five years and regress the yield changes of each bond on a constant and the first three principal components. The results for the data (based on 634 observations) and the two models (based on one million simulated observations) are shown in Table 6.

Panel A of Table 6 shows that the first three principal components describe almost all the variation of bond yield changes in the nonlinear model which is consistent with the data. Moreover, Panel B of Table 6 shows that the loading for each yield on the level, slope, and curvature factor in the nonlinear model is similar to the data. We conclude that the cross-sectional variation of bond yields implied by the nonlinear model is well explained by the first three principal components and no yield breaks this linear relation.

4 Conclusion

We introduce a new reduced form term structure model where the short rate and market prices of risk are nonlinear functions of Gaussian state variables and derive closed form solutions for yields. The nonlinear model with three Gaussian factors matches both the time-variation in expected excess returns and yield volatilities of U.S. Treasury bonds from 1961 to 2014. Because there are nonlinear relations between factors, yields, and variances, the model exhibits features consistent with empirical

evidence on unspanned risk premia (URP) and unspanned stochastic volatility (USV). We are not aware of any term structure models—in particular a model with only three factors—that have empirical properties consistent with evidence on time-variation in expected excess returns and volatilities, URP, and USV.

Although our empirical analysis has focused on a nonlinear generalization of an affine Gaussian model, it is possible to generalize a wide range of term structure models such as affine models with stochastic volatility and quadratic models. Our generalization introduces new dynamics for bond returns while keeping the new model as tractable as the standard model. Furthermore, the method extends to processes such as jump-diffusions and continuous time Markov chains.

References

- Ahn, D.-H, R.F. Dittmar, and A.R. Gallant, 2002, Quadratic term structure models: Theory and evidence, *Review of Financial Studies* 15, 243–288.
- Ahn, Dong-Hyun, Robert F. Dittmar, A. Ronald Gallant, and Bin Gao, 2003, Pure-bred or hybrid? Reproducing the volatility in term structure dynamics, *Journal of Econometrics* 116, 147–180.
- Ahn, Dong-Hyun, and Bin Gao, 1999, A parametric non-linear model of term structure dynamics, *Review of Financial Studies* 12, 721–762.
- Ait-Sahalia, Yacine, 1996a, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527–560.
- , 1996b, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

- Almeida, Caio, Jeremy J. Graveline, and Scott Joslin, 2011, Do interest rate options contain information about excess bond returns?, *Journal of Econometrics* 164, 35–44.
- Andersen, Torben Gustav, and Luca Benzoni, 2010, Do Bonds Span Volatility Risk in the U.S. Treasury market? A Specification Test for Affine Term Structure Models, *Journal of Finance* 65, 603–653.
- Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, 2010, Parametric and nonparametric volatility measurement, in Yacine Ait-Sahalia, and Lars Peter Hansen, ed.: *Handbook of Financial Econometrics, Vol 1: Tools and Techniques (Handbooks in Finance)* . pp. 67–137 (Elsevier).
- Ang, Andrew, and Geert Bekaert, 2002, Short rate nonlinearities and regime switches, *Journal of Economic Dynamics and Control* 26, 1243–1274.
- , and Min Wei, 2007, Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* 54, 1163–1212.
- Balduzzi, Pierluigi, and I-Hsuan Ethan Chiang, 2012, A simple test of the affine class of term structure models, *Review of Asset Pricing Studies* 2, 203–244.
- Bansal, Ravi, George Tauchen, and Hao Zhou, 2004, Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle, *Journal of Business and Economic Statistics* 22, 396–409.
- Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *Journal of Finance* 57, 1997–2043.
- Barndorff-Nielsen, Ole, and Neil Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B* 64, 253–280.

- Basak, Suleyman, 2000, A model of dynamic equilibrium asset pricing with heterogeneous beliefs and extraneous risk, *Journal of Economic Dynamics and Control* 24, 63–95.
- Bauer, Michael, and Glenn Rudebusch, 2015, Resolving the spanning puzzle in macro-finance term structure models, Working paper.
- Bhamra, Harjoat, and Raman Uppal, 2014, Asset prices with heterogeneity in preferences and beliefs, *Review of Financial Studies* 27, 519–580.
- Bikbov, R., and M. Chernov, 2009, Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options, *Management Science* pp. 1292–1305.
- Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *The Journal of Political Economy* 107, 205–251.
- Campbell, John Y., and Rober J. Shiller, 1991, Yield spread and interest rate movements: A bird’s eye view, *Review of Economic Studies* 58, 495–514.
- Carr, Peter, Xavier Gabaix, and Liuren Wu, 2009, Linearity-generating processes, unspanned stochastic volatility, and interest-rate option pricing, Working paper.
- Carr, Peter, and Liuren Wu, 2009, Stock options and credit default swaps: A joint framework for valuation amd estimation, *Journal of Financial Econometrics* pp. 1–41.
- Chan, K.C., Andrew Karolyi, Francis Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

- Chapman, David A., and Neil D. Pearson, 2000, Is the short rate drift actually nonlinear?, *Journal of Finance* LV, 355–388.
- Chen, Hui, and Scott Joslin, 2012, Generalized transform analysis of affine processes and applications in finance, *Review of Financial Studies* 25, 2225–2256.
- Chernov, Mikhail, and Philippe Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.
- Christoffersen, Peter, Christian Dorion, Kris Jacobs, and Lotfi Karoui, 2014, Nonlinear Kalman Filtering in Affine Term Structure Models, *CIRPEE Working Paper*.
- Cieslak, Anna, and Pavel Povala, 2014, Information in the term structure of yield curve volatility, forthcoming, *Journal of Finance*.
- , 2015, Expected returns in treasury bonds, conditionally accepted, *Review of Financial Studies*.
- Cochrane, John, and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.
- Cochrane, J. H., F. A. Longstaff, and P. Santa-Clara, 2008, Two trees, *Review of Financial Studies* 21, 347–385.
- Collin-Dufresne, Pierre, and Robert S. Goldstein, 2002, Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility, *Journal of Finance* 57, 1685–1730.
- , and Christopher S. Jones, 2009, Can interest rate volatility be extracted from the cross section of bond yields?, *Journal of Financial Economics* 94, 47–66.
- Constantinides, George M., 1992, A theory of the nominal term structure of interest rates, *Review of Financial Studies* 5, 531–552.

- Cooper, Ilan, and Richard Priestley, 2009, Time-Varying Risk Premiums and the Output Gap, *Review of Financial Studies* 22, 2601–2633.
- Cox, John. C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, p. 385–408.
- Creal, Drew D., and Jing Cynthia Wu, 2014, Estimation of affine term structure models with spanned or unspanned stochastic volatility, Working paper.
- , 2015, Estimation of affine term structure models with spanned or unspanned stochastic volatility, *Journal of Econometrics* 185, 60–81.
- Dai, Qiang, Anh Le, and Kenneth Singleton, 2010, Discrete-time affine^Q term structure models with generalized market prices of risk, *Review of Financial Studies* 23, 2184–2227.
- Dai, Q., and K. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943–1978.
- , 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631–678.
- Dai, Qiang, and Kenneth J. Singleton, 2002, Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63, 415–441.
- , and Wei Yang, 2007, Regime shifts in a dynamic term structure model of u.s. treasury bond yields, *Review of Financial Studies* 20, 1669–1706.
- Duffee, Gregory, 2002, Term premia and interest rate forecast in affine models, *Journal of Finance* 57, 405–443.
- , 2010, Sharpe ratios in term structure models, *Working Paper, University of California, Berkeley*.

- , 2011a, Forecasting with the term structure: The role of no-arbitrage restrictions, *Working Paper, Johns Hopkins University*.
- , 2011b, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895–2934.
- , 2012, Forecasting interest rates, Working paper.
- Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, p. 379–406.
- Dumas, Bernard, Alexander Kurshev, and Raman Uppal, 2009, Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility, *The Journal of Finance* 64, 579–629.
- Ehling, Paul, Michael Gallmeyer, Christian Heyerdahl-Larsen, and Philipp Illeditsch, 2013, Disagreement about inflation and the yield curve, Working Paper.
- Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.
- Fan, Rong, Anurag Gupta, and Peter Ritchken, 2003, Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets, *Journal of Finance* 58, 2219–2248.
- Feldhütter, Peter, 2008, Can affine models match the moments in bond yields?, working paper pp. 1–51.
- , Christian Heyerdahl-Larsen, and Philipp Illeditsch, 2015, Expanded term structure models, Work in progress.
- Feldhütter, Peter, and David Lando, 2008, Decomposing Swap Spreads, *JFE* 88, 375–405.

- Gabaix, Xavier, 2009, Linearity-generating processes: A modelling tool yielding closed forms for asset prices, CEPR and NBER Working paper.
- Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, The u.s. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291–2304.
- Heidari, Massoud, and Liuren Wu, 2003, Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?, *Journal of Fixed Income* 12, 75–86.
- Jacobs, Kris, and Lotfi Karoui, 2009, Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets, *Journal of Financial Economics* 91, 288–318.
- Jermann, Urban, 2013, A production-based model for the term structure, *Journal of Financial Economics* 109, 293–306.
- Jones, Christopher S., 2003, Nonlinear mean reversion in the short-term interest rate, *Review of Financial Studies* 16, 793–843.
- Joslin, Scott, 2014, Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?, *forthcoming, Management Science*.
- , Marcel Pribsch, and Kenneth Singleton, 2014, Risk premiums in dynamic term structure models with unspanned macro risks, *JF* 69, 1197–1233.
- Judd, Kenneth L., 1998, *Numerical Methods in Economics* (MIT Press) 1 edn.
- Leippold, Markus, and Liuren Wu, 2003, Design and Estimation of Quadratic Term Structure Models, *European Finance Review* 7, 47–73.
- Li, Haitao, and Feng Zhao, 2006, Unspanned stochastic volatility: Evidence from hedging interest rate derivatives, *Journal of Finance* 61, 341–378.

- Ludvigson, Sydney, and Serena Ng, 2009, Macro Factors in Bond Risk Premia, *Review of Financial Studies* 22, 5027–5067.
- Pritsker, Matt, 1998, Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies* 11, 449–487.
- Richard, Scott, 2013, A non-linear macroeconomic term structure model, Working Paper.
- Rogers, L.C.G., 1997, The potential approach to the term structure of interest rates and foreign exchange rates, *Mathematical Finance* 7, 157–164.
- Stanton, Richard, 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* VII, 1973–2002.
- Tang, Huarong, and Yihong Xia, 2007, An international examination of affine term structure models and the expectation hypothesis, *Journal of Financial and Quantitative Analysis* 42, 41–80.
- Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.
- Wright, Jonathan, 2011, Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset, *American Economic Review* 101, 1514–1534.
- Yan, Hongjun, 2008, Natural selection in financial markets: Does it work?, *Management Science* 54, 1935–1950.

A General Nonlinear Gaussian Model

Let γ denote a nonnegative constant and $M_0(t)$ a strictly positive stochastic process with dynamics given in equation (3). The stochastic discount factor is defined as

$$M(t) = M_0(t) \left(1 + \gamma e^{-\beta' X(t)}\right)^\alpha, \quad (38)$$

where $\beta \in \mathcal{R}^d$ and $\alpha \in \mathcal{N}$.

We show in the next theorem that the price of a bond is a weighted average of bond prices in artificial economies that belong to the class of essentially affine Gaussian term structure models.

Theorem 2. *The price of a zero-coupon bond that matures at time T is*

$$P(t, T) = \sum_{n=0}^{\alpha} s_n(t) P_n(t, T), \quad (39)$$

where

$$P_n(t, T) = e^{A_n(T-t) + B_n(T-t)' X(t)}, \quad (40)$$

$$s_n(t) = \frac{\binom{\alpha}{n} \gamma^n e^{-n\beta' X(t)}}{(1 + \gamma e^{-\beta' X(t)})^\alpha}. \quad (41)$$

The coefficient $A_n(T - t)$ and the d -dimensional vector $B_n(T - t)$ solve the ordinary differential equations given in equation (10) and (11).

Proof. Using the binomial expansion theorem, the stochastic discount factor in Equation (38) can be expanded as

$$M(t) = \sum_{n=0}^{\alpha} M_n(t), \quad (42)$$

where

$$M_n(t) = \binom{\alpha}{n} \gamma^n e^{-n\beta'X(t)} M_0(t). \quad (43)$$

Each summand can be interpreted as a stochastic discount factor in an artificial economy.²⁵ The dynamics of the strictly positive stochastic process $M_n(t)$ are

$$\frac{dM_n(t)}{M_n(t)} = -r_n(t) dt - \Lambda_n(t)' dW(t), \quad (44)$$

where

$$\Lambda_n(t) = \Lambda_0(t) + n\Sigma'\beta \quad (45)$$

$$r_n(t) = r_0(t) + n\beta'\kappa(\bar{X} - X(t)) - \frac{n^2}{2}\beta'\Sigma\Sigma'\beta - n\beta'\Sigma\Lambda_0(t). \quad (46)$$

Plugging in for $r_0(t)$ and $\Lambda_0(t)$, it is straightforward to show that $\Lambda_n(t)$ and $r_n(t)$ are affine functions of $X(t)$ with coefficients given in Equations (12)-(15). If $M_n(t)$ is interpreted as a stochastic discount factor of an artificial economy indexed by n then we know that bond prices in this economy belong to the class of essentially (exponential) affine Gaussian term structure models and hence

$$P_n(t, T) = e^{A_n(T-t) + B_n(T-t)'X(t)}, \quad (47)$$

where coefficient $A_n(T-t)$ and the d -dimensional vector $B_n(T-t)$ solve the ordinary differential equations (10) and (11). Hence, the bond price is

$$P(t, T) = \sum_{n=0}^{\alpha} s_n(t) P_n(t, T), \quad (48)$$

where $s_n(t)$ is given in equation (41).

²⁵Similar expansions of the stochastic discount factor appear in Yan (2008), Dumas, Kurshev, and Uppal (2009), Bhamra and Uppal (2014), and Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2013).

□

Proof of Theorem 1. Set $\alpha = 1$ in Theorem 2. □

Applying Ito's lemma to equation (38) leads to the dynamics of the stochastic discount factor:

$$\frac{dM(t)}{M(t)} = -r(t) dt - \Lambda(t)' dW(t), \quad (49)$$

where

$$\begin{aligned} r(t) = & r_0(t) + \alpha(1 - s(t))\beta' \kappa (\bar{X} - X(t)) - \alpha(1 - s(t))\beta' \Sigma \Lambda_0(t) \\ & - \frac{\alpha}{2}(1 - s(t))(\alpha(1 - s(t)) + s(t))\beta' \Sigma \Sigma' \beta. \end{aligned} \quad (50)$$

and

$$\Lambda(t) = \Lambda_0(t) + \alpha(1 - s(t))\Sigma' \beta. \quad (51)$$

Let $\omega_n(t, T)$ denote the contribution of each artificial exponential affine bond price to the total bond price. Specifically,

$$\omega_n(t, T) = \frac{P_n(t, T)s_n(t)}{P(t, T)}. \quad (52)$$

The dynamics of the bond price $P(t, T)$ are

$$\frac{dP(t, T)}{P(t, T)} = (r(t) + \Lambda(t)'\sigma(t, T)) dt + \sigma(t, T)' dW(t), \quad (53)$$

where

$$\sigma(t, T) = \Sigma' \left(\sum_{n=0}^{\alpha} \omega_n(t, T) B_n(T - t) + \beta \left(\sum_{n=0}^{\alpha} n \omega_n(t, T) - \alpha(1 - s(t)) \right) \right). \quad (54)$$

B Essentially Affine $A_1(3)$ Model

We briefly describe the $A_1(3)$ model in this section and refer the reader to Duffee (2002) for a detailed discussion. The dynamics of the three-dimensional state vector $X(t) = (X_1(t), X_2(t), X_3(t))'$ are

$$dX(t) = \kappa (\bar{X} - X(t)) dt + S(t) dW(t), \quad (55)$$

where $\bar{X} = (\bar{X}_1, 0, 0)'$ is the long run mean,

$$\kappa = \begin{pmatrix} \kappa_{(1,1)} & 0 & 0 \\ \kappa_{(2,1)} & \kappa_{(2,2)} & \kappa_{(2,3)} \\ \kappa_{(3,1)} & \kappa_{(3,2)} & \kappa_{(3,3)} \end{pmatrix} \quad (56)$$

is the positive-definite mean reversion matrix, $W(t)$ is a three-dimensional Brownian motion, and

$$S(t) = \begin{pmatrix} \sqrt{\delta_1 X_1(t)} & 0 & 0 \\ 0 & \sqrt{1 + \delta_2 X_1(t)} & 0 \\ 0 & 0 & \sqrt{1 + \delta_3 X_1(t)} \end{pmatrix} \quad (57)$$

is the local volatility matrix with $\delta = (1, \delta_2, \delta_3)$.

The dynamics of the stochastic discount factor $M(t)$ are

$$\frac{dM(t)}{M(t)} = -r(t) dt - \Lambda(t)' dW(t), \quad (58)$$

where the short rate $r(t)$ and the three-dimensional vector $S(t)\Lambda(t)$ are affine functions of $X(t)$. Specifically,

$$r(t) = \rho_0 + \rho_X' X(t), \quad (59)$$

where ρ_0 is a scalar and ρ_X is a 3-dimensional vector. The market price of risk $\Lambda(t)$ is the solution of the equation

$$S(t)\Lambda(t) = \begin{pmatrix} \lambda_{X,(1,1)}X_1(t) \\ \lambda_{0,2} + \lambda_{X,(2,1)}X_1(t) + \lambda_{X,(2,2)}X_2(t) + \lambda_{X,(2,3)}X_3(t) \\ \lambda_{0,3} + \lambda_{X,(3,1)}X_1(t) + \lambda_{X,(3,2)}X_2(t) + \lambda_{X,(3,3)}X_3(t) \end{pmatrix}, \quad (60)$$

where λ_0 denotes a three dimensional vector and λ_X a three-dimensional matrix.

Hence, the bond price and the instantaneous yield volatility are

$$P(X(t), T) = e^{A(T-t) + B(T-t)'X(t)} \quad (61)$$

$$v(X(t), T) = \sqrt{B(T-t)'S(X(t))S(X(t))B(T-t)}, \quad (62)$$

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

$$\frac{dA(\tau)}{d\tau} = (\kappa\bar{X} - \lambda_0)' B(\tau) + \frac{1}{2} \sum_{i=2}^3 B_i(\tau)^2 - \rho_0, \quad A(0) = 0 \quad (63)$$

$$\frac{dB(\tau)}{d\tau} = (\kappa + \lambda_X)' B(\tau) + \frac{1}{2} \sum_{i=1}^3 B_i(\tau)\delta_i - \rho_X, \quad B(0) = 0_{3 \times 1}. \quad (64)$$

C Equilibrium Models

In this section we show that the functional form of the state price density in equation (2) and (38) naturally comes out of several equilibrium models.²⁶ We need to allow for state variables that follow arithmetic Brownian motions and hence we rewrite the dynamics of the state vector in equation (1) in the slightly more general form

$$dX(t) = (\theta - \kappa X(t)) dt + \Sigma dW(t), \quad (65)$$

²⁶Chen and Joslin (2012) provide an alternative way to solve many of these equilibrium models that is based on a nonlinear transform of processes with tractable characteristic functions.

where θ is d -dimensional and κ and Σ are $d \times d$ -dimensional.

In what follows, the standard consumption based asset pricing model with a representative agent power utility and log-normally distributed consumption will serve as our benchmark model. Specifically, the state price density takes the following form

$$M_0(t) = e^{-\rho t} C(t)^{-R}, \quad (66)$$

where R is the coefficient of RRA and C is aggregate consumption with dynamics

$$\frac{dC(t)}{C(t)} = \mu_C dt + \sigma'_C dW(t). \quad (67)$$

The short rate and the market price of risk are both constant and given by

$$\Lambda_0 = R\sigma_C \quad (68)$$

$$r_0 = \rho + R\mu_C - \frac{1}{2}R(R+1)\sigma'_C\sigma_C. \quad (69)$$

Table 7 summarizes the relation between the nonlinear term structure models and the equilibrium models discussed in this section.

C.1 Two Trees

Cochrane, Longstaff, and Santa-Clara (2008) study an economy in which aggregate consumption is the sum of two Lucas trees. In particular they assume that the dividends of each tree follow a geometric Brownian motion

$$dD_i(t) = D_i(t) (\mu_i dt + \sigma'_i dW(t)). \quad (70)$$

Aggregate consumption is $C(t) = D_1(t) + D_2(t)$. There is a representative agent with

power utility and risk aversion R . Hence, the stochastic discount factor is

$$\begin{aligned}
M(t) &= e^{-\rho t} C(t)^{-R} \\
&= e^{-\rho t} (D_1(t) + D_2(t))^{-R} \\
&= e^{-\rho t} D_1(t)^{-R} \left(1 + \frac{D_2(t)}{D_1(t)} \right)^{-R} \\
&= M_0(t) \left(1 + e^{\log(D_2(t)) - \log(D_1(t))} \right)^{-R}, \tag{71}
\end{aligned}$$

where $M_0(t) = e^{-\rho t} D_1(t)^{-R}$ and $X(t) = \log(D_1(t)/D_2(t))$. Equation (71) has the same form as the SDF in equation (38) with $\alpha \notin \mathcal{N}$. Specifically, $\gamma = 1$, $\beta = 1$, and $\alpha = -R$. Note that in this case the state variable is the log-ratio of two geometric Brownian motions and thus $\kappa = 0$. The share $s(X(t))$ and hence yields are not stationary.

C.2 Multiple Consumption Goods

Models with multiple consumption goods and CES consumption aggregator naturally falls within the functional form of the SDF in equation (38). Consider a setting with two consumption goods. The aggregate output of the two goods are given by

$$dD_i(t) = D_i(t) (\mu_i dt + \sigma'_i dW(t)). \tag{72}$$

Assume that the representative agent has the following utility over aggregate consumption C ,

$$u(C, t) = e^{-\rho t} \frac{1}{1-R} C^{1-R}, \tag{73}$$

where

$$C(C_1, C_2) = \left(\phi^{1-b} C_1^b + (1-\phi)^{1-b} C_2^b \right)^{\frac{1}{b}}. \tag{74}$$

We use the aggregate consumption bundle as numeraire, and consequently the state price density is

$$\begin{aligned}
M(t) &= e^{-\rho t} C(t)^{-R} \\
&= (\phi)^{\frac{bR}{1-b}} e^{-\rho t} D_1(t)^{-R} \left(1 + \left(\frac{1-\phi}{\phi} \right)^{1-b} \left(\frac{D_2(t)}{D_1(t)} \right)^b \right)^{-\frac{R}{b}}.
\end{aligned} \tag{75}$$

After normalizing equation (75) has the same form as the SDF in equation (38) with $\alpha \notin \mathcal{N}$. Specifically, $X(t) = \log(D_1(t)/D_2(t))$, $\gamma = \left(\frac{1-\phi}{\phi} \right)^{1-b}$, $\beta = b$, and $\alpha = -\frac{R}{b}$. As in the case with Two Trees, the share $s(X(t))$ and hence yields are not stationary.

C.3 External Habit Formation

The utility function in Campbell and Cochrane (1999) is

$$U(C, H) = e^{-\rho t} \frac{1}{1-R} (C - H)^{1-R}, \tag{76}$$

where H is the habit level. Rather than working directly with the habit level, Campbell and Cochrane (1999) define the surplus consumption ratio $s = \frac{C-H}{C}$. The stochastic discount factor is

$$M(t) = e^{-\rho t} C(t)^{-R} s(t)^{-R} \tag{77}$$

$$= M_0(t) s(t)^{-R}. \tag{78}$$

Define the state variable

$$dX(t) = \kappa (\bar{X} - X(t)) dt + b dW(t), \tag{79}$$

where $\kappa > 0, \sigma_c > 0$ and $b > 0$. Now let $s(t) = \frac{1}{1+e^{-\beta X(t)}}$. Note that $s(t)$ is between 0 and 1. In particular, $s(t)$ follows

$$ds(t) = s(t) (\mu_s(t)dt + \sigma_s(t)dW(t)), \quad (80)$$

where

$$\mu_s(t) = (1 - s(t)) \left(\beta \kappa (\bar{X} - X(t)) + \frac{1}{2} (1 - 2s(t)) \beta^2 b^2 \right) \quad (81)$$

$$\sigma_s(t) = (1 - s(t)) \beta b. \quad (82)$$

The functional form of the surplus consumption ratio differs from Campbell and Cochrane (1999). However, note that the surplus consumption ratio is locally perfectly correlated with consumption shocks, mean-reverting and bounded between 0 and 1 just as in Campbell and Cochrane (1999). The state price density can be written as

$$M(t) = M_0(t) (1 + e^{-\beta X(t)})^R. \quad (83)$$

The above state price density has the same form as equation (38) with parameters $\gamma = 1$, $\beta = \beta$, and $\alpha = R$. Note that the state variable X in this case is mean-reverting and therefore the share $s(X(t))$ and hence yields are stationary.

C.4 Heterogeneous Beliefs

Consider an economy with two agents that have different beliefs. Let both agents have power utility with the same coefficient of relative risk aversion, R . Moreover, assume that aggregate consumption follows the dynamics in equation (67). The agents do

not observe the expected growth rate and agree to disagree.²⁷ The equilibrium can be solved by forming the central planner problem with stochastic weight λ that captures the agents' initial relative wealth and their differences in beliefs (see Basak (2000), for example),

$$U(C, \lambda) = \max_{\{C_1+C_2=C\}} \left(\frac{1}{1-R} C_1^{1-R} + \lambda \frac{1}{1-R} C_2^{1-R} \right). \quad (84)$$

Solving the above problem leads to the optimal consumption of the agents

$$C_1(t) = s(t)C(t), \quad (85)$$

$$C_2(t) = (1 - s(t))C(t), \quad (86)$$

where $s(t) = \frac{1}{1+\lambda(t)^{\frac{1}{R}}}$ is the consumption share of the first agent and C is the aggregate consumption. The state price density as perceived by the first agent is

$$\begin{aligned} M(t) &= e^{-\rho t} C_1(t)^{-R} \\ &= e^{-\rho t} C(t)^{-R} s(t)^{-R} \\ &= M_0(t) \left(1 + e^{\frac{1}{R} \log(\lambda(t))} \right)^R. \end{aligned} \quad (87)$$

This has the same form as equation (38) with $X(t) = \log(\lambda(t))$, $\gamma = 1$, $\beta = -\frac{1}{R}$, and $\alpha = R$. The dynamics of the state variable is driven by the log-likelihood ratio of the two agents and consequently the share $s(X(t))$ and hence yields are not stationary.

²⁷The model can easily be generalised to a setting with disagreement about multiple stochastic processes and learning. For instance, Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2013) show that in a model with disagreement about inflation, the bond prices are weighted averages of quadratic Gaussian term structure models.

C.5 HARA Utility

Consider a pure exchange economy with a representative agent with utility $u(t, c) = \frac{e^{-\rho t}}{1-R} (C + b)^{1-R}$, where $R > 0$ and $b > 0$. We can write the SDF as

$$\begin{aligned}
 M(t) &= e^{-\rho t} C(t)^{-R} \\
 &= e^{-\rho t} (C(t) + b)^{-R} \\
 &= e^{-\rho t} C(t)^{-R} \left(1 + \frac{b}{C(t)}\right)^{-R} \\
 &= M_0(t) \left(1 + e^{\log(b) - \log(C(t))}\right)^{-R}
 \end{aligned} \tag{88}$$

After normalizing equation (88) has the same form as the SDF in equation (38) with $\alpha \notin \mathcal{N}$. Specifically, $X(t) = \log(b/C(t))$, $\gamma = 1$, $\beta = 1$, and $\alpha = -R$. Similarly to the model with Two Trees and multiple consumption goods, the share $s(X(t))$ and hence yields are nonstationary as the ratio $b/C(t)$ will eventually converge to zero or infinity depending on the expected growth in the economy.

D Gauss-Hermite Quadrature

While bond prices and bond yields are given in closed form, conditional moments of yields and bond returns are not. However, it is straightforward to calculate conditional expectations using Gauss-Hermite polynomials because the state vector $X(t)$ is Gaussian.²⁸

In this section we illustrate how to calculate the expectation of a function of Gaussian state variables. Let μ_X and Σ_X denote the conditional mean and variance of $X(u)$ at time $t < u$. Let $f(X(t))$ be a function of the state vector at time t . For instance if you want to calculate at time t the n -th uncentered moment of the

²⁸For more details see Judd (1998).

bond yield with maturity τ at time u , then $f(X(u)) = (y^{(\tau)}(X(u)))^n$. Hence, the conditional expectation of $y^{(\tau)}(X(u))$ at time t is

$$E_t[f(X(u))] = \int_{\mathcal{R}^d} f(x) \frac{1}{((2\pi)^d |\Sigma_X|)^{0.5}} e^{-\frac{1}{2}(x-\mu_X)' \Sigma_X^{-1} (x-\mu_X)} dx. \quad (89)$$

Define $y = \sqrt{2}\sigma_X^{-1}(x - \mu_X)$ where σ_X is determined by the Cholesky decomposition $\Sigma_X = \sigma_X \sigma_X'$. Hence, we can write Equation (89) as

$$\pi^{-\frac{d}{2}} \int_{\mathcal{R}^d} f(\sqrt{2}\sigma_X y + \mu_X) e^{-y'y} dy. \quad (90)$$

Let $g(y) = f(\sqrt{2}\sigma_X y + \mu_X)$. We set $d = 3$ in the empirical section of the paper and thus the integral in Equation (90) can be approximated by the n point Gauss-Hermite quadrature

$$\int_{\mathcal{R}^d} f(\sqrt{2}\sigma_X y + \mu_X) e^{-y'y} dy \approx \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_i w_j w_k g(y_1(i), y_2(j), y_3(k)), \quad (91)$$

where w_i are the weights and $y_l(i)$ are the nodes for the n point Gauss-Hermite quadrature for $i = 1, \dots, n$ and $l = 1, \dots, 3$. We use $n = 4$ in equation (91).

	Nonlinear			Nonlinear, post-Voelcker			$A_1(3)$		
κ	0.3127 (0.04224)	0	0	0.3452 (0.08753)	0	0	1.421 (0.1863)	0	0
	0.3063 (0.05601)	0.002189 (2.246e−05)	0	0.5507 (0.09825)	0.003245 (0.002091)	0	−0.04787 (1.899)	0.07225 (0.01938)	−0.003283 (4.101)
	1.258 (0.1103)	0.03804 (0.02125)	0.4098 (0.0377)	1.057 (0.2745)	1.072e − 05 (0.0002734)	0.4449 (0.2494)	0.283 (0.6523)	−0.009014 (0.07474)	0.356 (0.01893)
ρ_0	−0.001756 (0.01408)			−0.001002 (0.02238)			0.08832 (0.3038)		
ρ_X	0.0002071 (0.0001846)	0.003061 (0.0002364)	0.004345 (0.0001742)	0.0002036 (0.0009384)	0.005161 (0.0004481)	0.004939 (0.0005533)	0.0003736 (0.0002645)	0.001131 (0.0009603)	1.385e − 05 (0.000302)
λ_0	0.7569 (0.04302)	−0.01631 (0.5559)	−0.4413 (0.3375)	0.3814 (0.09227)	−0.02483 (0.09312)	−0.3191 (0.2209)	0	0.6101 (106.4)	0.006454 (7.178)
λ_X	−0.2187 (0.04129)	0.005572 (0.001321)	−0.02053 (0.005609)	−0.2244 (0.06907)	0.003604 (0.00792)	−0.02491 (0.04552)	6.75e − 05 (0.07544)	0	0
	−1.735e − 06 (4.238e−05)	0.001197 (0.03785)	0.6863 (0.03001)	−1.558e − 06 (2.248e−05)	0.001282 (0.03908)	0.7165 (0.05695)	2.378 (3.64)	−0.0006549 (0.01964)	3.381 (5.878)
	−0.2943 (0.1053)	−0.02387 (0.01562)	0.04613 (0.05121)	−0.3973 (0.2578)	−0.0237 (0.02542)	0.05947 (0.2159)	0.01683 (0.7003)	−0.0001671 (0.0733)	1.302e − 05 (0.01966)
γ	0.0003857 (0.0004591)			0.0005653 (0.0007368)					
β	−1.444 (0.008187)	−0.2376 (0.01831)	0.2846 (0.02526)	−1.196 (0.0521)	−0.2737 (0.07188)	0.3483 (0.08285)			
δ							491.5 (836.6) 2.417 (0.3336)		
$(\kappa X)_1$							1.509 (0.1109)		
σ_y	0.0005463 (6.945e−05)			0.0004679 (9.47e−05)			0.0006001 (8.676e−05)		
σ_{rv}	7.281e − 05 (8.491e−06)			2.857e − 05 (3.381e−06)			6.18e − 05 (6.019e−06)		

Table 1: *Parameter estimates.* This table contains parameter estimates and asymptotic standard errors (in parenthesis). The first and third columns show the parameters estimates based on an estimation where yield and realized variance data from the period 1961:07-2014:04 is used. The second column shows parameter estimates for the nonlinear model based on an estimation where data for the period 1987:08-2014:04 is used.

Panel A: Campbell-Shiller regression coefficients									
		bond maturity		2-year	3-year	4-year	5-year		
Data				-0.63 (0.64)	-0.93 (0.69)	-1.21 (0.73)	-1.47 (0.77)		
Nonlinear model				-0.61	-0.61	-0.63	-0.65		
$A_1(3)$ model				-0.01	0.01	0.04	0.07		
Panel B: Regressing realized excess returns on expected excess returns									
Maturity	Nonlinear model				$A_1(3)$ model				Cochrane-Piazzesi R^2
	$\alpha \times 10^3$	β	R^2	FVE	$\alpha \times 10^3$	β	R^2	FVE	
One-year holding horizon									
$\tau=2$	-2.58 (2.89)	0.81 (0.20)	0.22	0.15	2.53 (2.62)	0.96 (0.35)	0.12	0.10	0.15
$\tau=3$	-4.85 (5.06)	0.83 (0.20)	0.23	0.16	4.05 (4.64)	0.97 (0.33)	0.13	0.12	0.16
$\tau=4$	-6.60 (6.74)	0.85 (0.19)	0.24	0.18	5.07 (6.32)	1.00 (0.32)	0.15	0.13	0.18
$\tau=5$	-7.51 (8.09)	0.86 (0.19)	0.25	0.20	5.85 (7.77)	1.03 (0.32)	0.16	0.15	0.20
Two-year holding horizon									
$\tau=3$	-3.73 (4.60)	0.87 (0.22)	0.29	0.23	6.31 (4.73)	0.62 (0.38)	0.07	0.00	0.10
$\tau=4$	-6.36 (8.08)	0.89 (0.21)	0.30	0.25	10.75 (8.51)	0.67 (0.37)	0.08	0.03	0.11
$\tau=5$	-7.72 (10.70)	0.91 (0.21)	0.32	0.29	14.24 (11.71)	0.73 (0.36)	0.10	0.05	0.13
Three-year holding horizon									
$\tau=4$	-2.30 (5.54)	0.83 (0.23)	0.29	0.24	10.30 (6.45)	0.31 (0.41)	0.02	-0.15	0.13
$\tau=5$	-3.96 (10.05)	0.87 (0.22)	0.33	0.29	17.60 (12.09)	0.40 (0.40)	0.03	-0.11	0.15
Four-year holding horizon									
$\tau=5$	0.53 (6.57)	0.75 (0.24)	0.25	0.20	12.72 (8.12)	0.25 (0.42)	0.01	-0.21	0.23

Table 2: *Excess return regressions*. Panel A shows the coefficients ϕ^τ from the regressions $y(t+1, \tau-1) - y(t, \tau) = \text{const} + \phi^\tau \left[\frac{y(t, \tau) - y(t, 1)}{\tau-1} \right] + \text{residual}$, where $y(t, \tau)$ is the zero-coupon yield at time t of a bond maturing at time $t + \tau$ (τ and t are measured in years). The actual coefficients are calculated using monthly data of one through five-year zero coupon bond yields from 1961:7 to 2014:04 obtained from Gurkaynak, Sack, and Wright (2007). For each model the coefficient is based on one simulated sample path of 1,000,000 months. Panel B shows regression coefficients from a regression of realized (log) excess returns on expected (log) excess returns in sample. The FVE is $1 - \frac{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - E_t[rx_{t,t+n}^\tau])^2}{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - \frac{1}{T} \sum_{t=1}^T rx_{t,t+n}^\tau)^2}$, where $rx_{t,t+n}^\tau$ is the n -year excess return on a bond with maturity τ and $E_t[rx_{t,t+n}^\tau]$ is the corresponding model implied expected excess return. The last row contains the R^2 from the regression of future realized excess returns on the five yields. For both panels standard errors in parentheses are computed using the Hansen and Hodrick (1980) correction with number of lags equal to the number of overlapping months.

Maturity	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1-PC_4	PC_1-PC_5	Full
Panel A: R^2 in data						
$\tau = 2$	2.1	12.6	13.2	14.4	14.6	
$\tau = 3$	0.8	13.9	14.3	15.9	16.2	
$\tau = 4$	0.3	15.6	15.8	17.8	18.1	
$\tau = 5$	0.1	17.2	17.3	19.7	19.9	
Panel B: R^2 for nonlinear model in data sample						
$\tau = 2$	7.8	64.7	66.9	85.2	91.1	100.0
$\tau = 3$	6.8	67.5	68.6	84.8	90.8	100.0
$\tau = 4$	6.5	69.5	70.3	84.9	90.9	100.0
$\tau = 5$	6.9	70.8	71.7	85.6	91.2	100.0
Panel C: R^2 for $A_1(3)$ model in data sample						
$\tau = 2$	10.8	99.8	100.0			100.0
$\tau = 3$	10.5	99.7	100.0			100.0
$\tau = 4$	10.2	99.6	100.0			100.0
$\tau = 5$	9.9	99.5	100.0			100.0
Panel D: R^2 for nonlinear model in population						
$\tau = 2$	0.0	10.7	14.5	15.7	15.8	28.0
$\tau = 3$	0.1	10.5	14.4	15.2	15.3	26.2
$\tau = 4$	0.1	10.6	14.5	15.2	15.3	25.3
$\tau = 5$	0.1	11.1	14.7	15.6	15.6	25.2
Panel E: R^2 for $A_1(3)$ model in population						
$\tau = 2$	3.9	4.5	4.5	4.5	4.5	4.5
$\tau = 3$	3.9	4.5	4.5	4.5	4.5	4.5
$\tau = 4$	3.9	4.5	4.5	4.5	4.5	4.5
$\tau = 5$	3.9	4.5	4.5	4.5	4.5	4.5

Table 3: *URP (Unspanned Risk Premia) regressions*. This table shows R^2 s (in percent) from regressions of excess returns (measured monthly) on the five principal components (PCs) of yields. The final column shows—where relevant—the R^2 s when using model-implied excess return on the righthandside instead of the model-implied PCs. Panel A shows one-year actual realized excess return on PCs of actual yields in the data sample 1961:07-2014:04. Panel B shows for the nonlinear model model-implied one-year excess return on model-implied PCs of yields, 1961:07-2014:04. Panel C shows for the $A_1(3)$ model model-implied one-year excess return on model-implied PCs of yields, 1961:07-2014:04. Panel D and E shows for the nonlinear respectively $A_1(3)$ model R^2 s in a regression of realized one-year excess return on PCs of yields in a simulated data sample of 1,000,000 months.

Maturity	PC_1-PC_3			PC_1-PC_4			PC_1-PC_5		
	R^2	slope	t -stat.	R^2	slope	t -stat.	R^2	slope	t -stat.
$\tau = 2$	11.52	-0.0011	-1.97	20.56	-0.0011	-4.27	17.76	-0.0008	-3.88
$\tau = 3$	11.66	-0.0020	-2.09	19.55	-0.0020	-4.00	16.44	-0.0015	-3.63
$\tau = 4$	11.56	-0.0027	-2.17	18.65	-0.0027	-3.79	15.28	-0.0020	-3.39
$\tau = 5$	11.09	-0.0033	-2.15	17.78	-0.0032	-3.59	14.17	-0.0024	-3.17

Table 4: *URP (Unspanned Risk Premia) and expected inflation.* This table shows R^2 s (in percent), slope and the t -statistic. We use the mean forecasts of the Michigan Survey of Consumers (MSC) to measure expected inflation. The residual from the URP regressions are from regressing the model implied expected excess return on the first three PCs (PC_1-PC_3), the first four PCs (PC_1-PC_4) and all five PCs (PC_1-PC_5). Model implied excess returns are measured as the expected one year return in excess of the one year yield. Standard errors are Newey-West corrected using 12 lags. The data sample is 1978:1-2014:4 as MSC is only available in monthly frequencies starting in 1978.

Maturity	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1-PC_4	PC_1-PC_5
Panel A: R^2 for data (realized variance, monthly observations, 1961-2014)					
$\tau = 1$	24.3	26.8	35.0	35.7	40.2
$\tau = 2$	23.2	24.8	33.7	35.4	41.6
$\tau = 3$	21.9	22.8	32.6	35.8	42.5
$\tau = 4$	20.3	20.7	31.1	35.9	42.6
$\tau = 5$	18.8	18.9	29.6	36.0	42.6
Panel B: R^2 for nonlinear model (instantaneous variance, 1961-2014)					
$\tau = 1$	21.6	21.8	44.0	47.9	55.1
$\tau = 2$	19.1	19.1	42.3	49.2	57.4
$\tau = 3$	17.5	17.6	41.8	50.9	59.9
$\tau = 4$	16.7	16.8	42.0	52.9	61.7
$\tau = 5$	16.9	17.2	42.4	54.6	62.1
Panel C: R^2 for nonlinear model (realized variance, 1,000,000 simulated months)					
$\tau = 1$	31.8	32.7	40.8	46.0	56.9
$\tau = 2$	32.8	33.8	40.7	48.6	60.8
$\tau = 3$	32.9	34.2	40.1	50.3	63.4
$\tau = 4$	32.6	34.4	39.2	51.0	65.0
$\tau = 5$	31.9	34.7	38.3	51.0	66.1
Panel D: R^2 for $A_1(3)$ model (instantaneous variance, 1961-2014)					
$\tau = 1, \dots, 5$	21.5	22.3	100.0	100.0	100.0
Panel E: R^2 for $A_1(3)$ model (realized variance, 1,000,000 simulated months)					
$\tau = 1, \dots, 5$	0.0	0.0	45.8	45.8	45.8
Panel F: Percent of residual variation explained by residual PCs, 1961-2014					
Data	91.8	98.7	99.9	100.0	100.0
Nonlinear model	97.9	99.9	100.0	100.0	100.0

Table 5: *USV (Unspanned Stochastic Volatility) regressions*. Panel A shows R^2 s (in percent) from regressing realized variance on the five principal components (PCs) of yields. Panel B shows R^2 s for the nonlinear model from regressing model-implied instantaneous variance on the PCs of model-implied yields in the sample period. Panel C shows R^2 s for the nonlinear model from regressing monthly realized variance (based on daily model-implied yields) on the PCs of monthly yields (based on averages over daily model-implied yields) in a sample of 1,000,000 simulated months. Panel D and E shows corresponding results for the $A_1(3)$ model corresponding to Panel B and C for the nonlinear model, where only results for one maturity is shown because R^2 's are the same for all maturities. Panel F shows the explanatory power of the PCs of the residuals of the regressions in Panel A and B.

Panel A:	PCA of yield changes			PCA of realized variances		
	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1	PC_1-PC_2	PC_1-PC_3
Data	0.9550	0.9975	0.9999	0.9454	0.9922	0.9996
Nonlinear	0.9819	0.9995	1.0000	0.9750	0.9993	1.0000
$A_1(3)$	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000

Panel B:	Yield changes on PCs			Realized variances on PCs		
	PC_1	PC_2	PC_3	PC_1	PC_2	PC_3
Maturity	Data					
$\tau = 1$	0.47	0.72	0.48	0.50	0.74	-0.43
$\tau = 2$	0.46	0.22	-0.52	0.49	0.14	0.60
$\tau = 3$	0.45	-0.12	-0.46	0.46	-0.22	0.38
$\tau = 4$	0.43	-0.36	-0.02	0.41	-0.41	-0.12
$\tau = 5$	0.42	-0.54	0.54	0.37	-0.47	-0.54
	Nonlinear Model					
$\tau = 1$	0.46	0.67	0.52	0.53	0.64	-0.48
$\tau = 2$	0.45	0.28	-0.36	0.49	0.20	0.40
$\tau = 3$	0.45	-0.05	-0.52	0.44	-0.16	0.50
$\tau = 4$	0.44	-0.34	-0.17	0.39	-0.42	0.09
$\tau = 5$	0.44	-0.60	0.54	0.34	-0.59	-0.59
	$A_1(3)$ Model					
$\tau = 1$	0.52	0.66	-0.53	0.54		
$\tau = 2$	0.48	0.21	0.58	0.49		
$\tau = 3$	0.44	-0.14	0.41	0.44		
$\tau = 4$	0.41	-0.39	-0.04	0.39		
$\tau = 5$	0.38	-0.58	-0.46	0.35		

Table 6: *Cross-sectional fit of three-factor models.* Principal components are constructed from a panel of constant-maturity zero-coupon bond yield changes and from a panel of realized variances of constant-maturity zero-coupon bond yields. Maturities are ranging from one to five years in both panels. Panel A shows the contribution of the first three principal components to the total variation in bond yield changes (columns 2-4) and realized yield variances (columns 5-7). Columns 2-4 of Panel B show the slope coefficients from the regressions of each yield on a constant and the first three principal components of yield changes. Columns 5-7 of Panel B show the slope coefficients from the regressions of each realized yield variance on a constant and the first three principal components of realized yield variances. The actual coefficients are computed using monthly data of one through five-year zero coupon bond yield changes and their realized variances from 1961:07 to 2014:04. For each model the coefficient is based on one simulated sample path of 1,000,000 months and the monthly realized variances are based on squared changes of simulated yields.

Model	N	d	X	α	γ	β	Stationary
Two trees	1	2	$\log(D_1(t)/D_2(t))$	$-R$	1	1	No
Multiple consumption goods	1	2	$\log(D_1(t)/D_2(t))$	$-\frac{R}{b}$	$\left(\frac{1-\phi}{\phi}\right)^{1-b}$	b	No
External habit formation	1	1	X	R	1	β	Yes
Heterogeneous beliefs	1	1	$\log(\lambda(t))$	R	1	$-\frac{1}{R}$	No
HARA utility	1	1	$\log(b/C(t))$	$-R$	1	1	No

Table 7: *Equilibrium models.* The table shows various equilibrium models and how they map into the nonlinear term structure models.

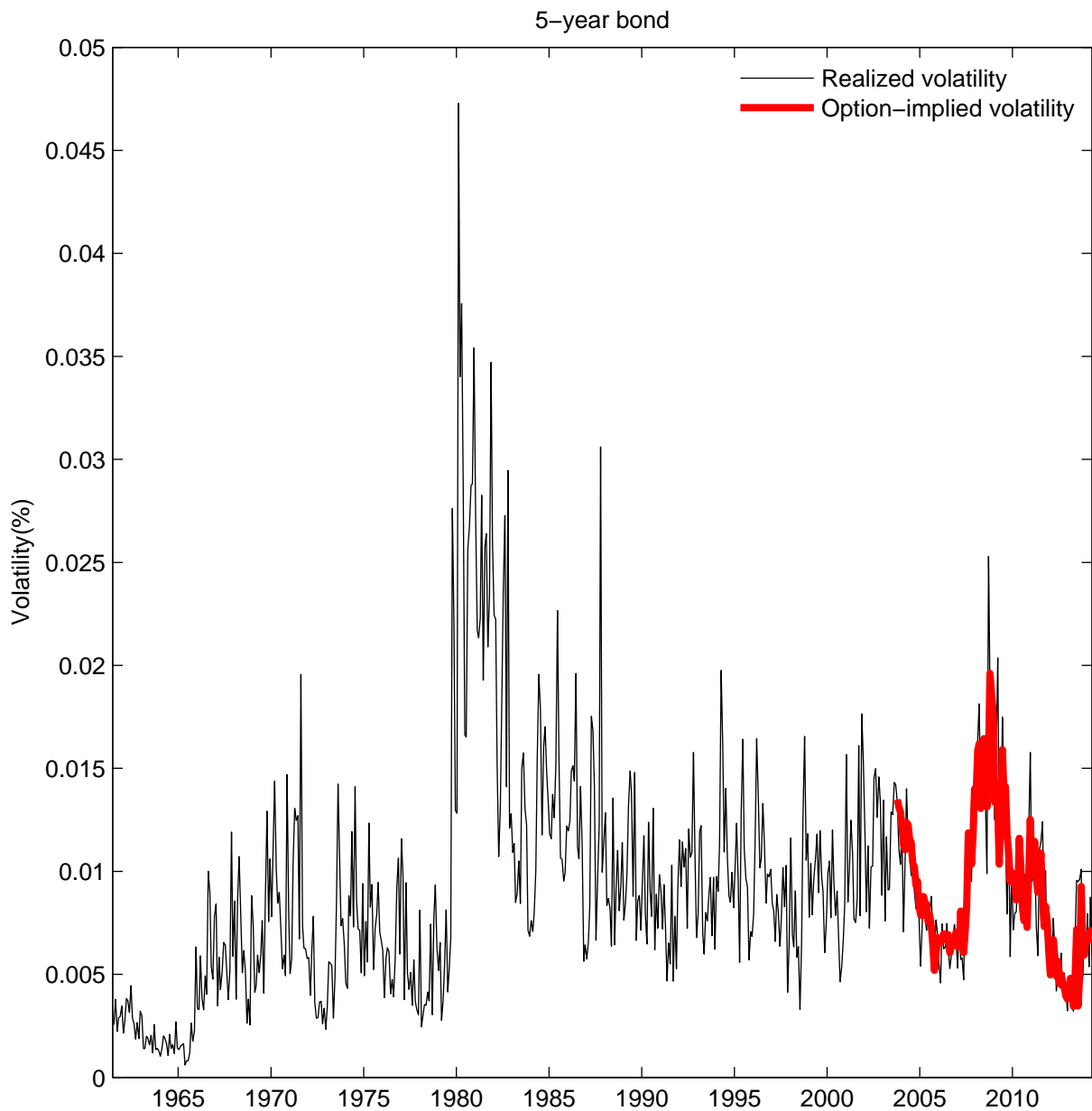


Figure 1: *Realized and option-implied yield volatility.* We use monthly estimates of realized yield variance based on daily squared yield changes. This graph shows that option-implied volatility tracks the realized volatility closely over the last 10 years (the correlation is 87%). Option-implied volatility is obtained from 1-month at-the-money options on 5-year Treasury futures as explained in the text. The data are available from Datastream since October 2003.

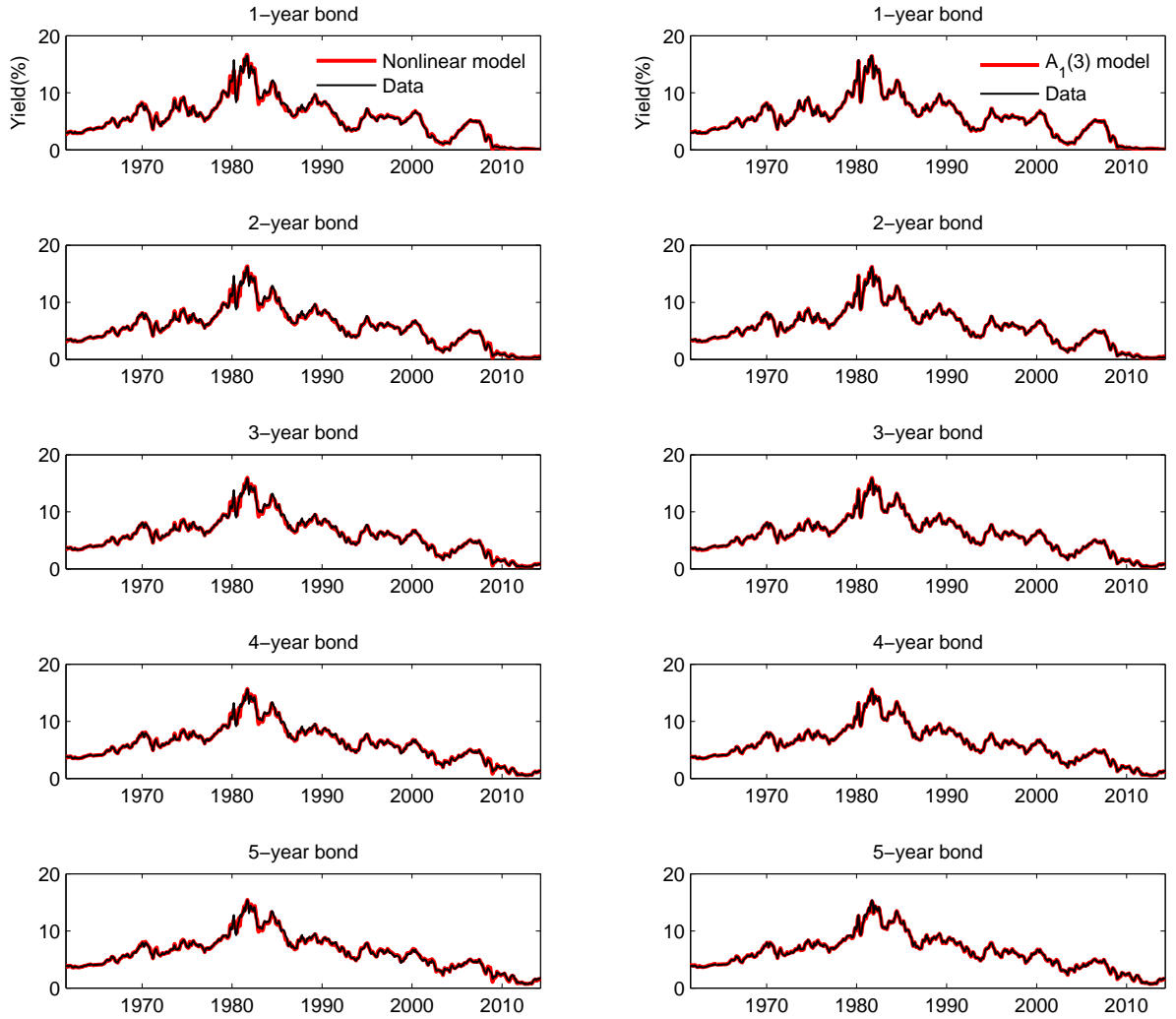


Figure 2: *Fit to yields*. This figure shows the actual yield along the estimated yield in the nonlinear and the $A_1(3)$ model. The actual yield is calculated on a monthly basis and is based on the average daily yield over the month. We use daily data from Gurkaynak, Sack, and Wright(2007) for the period 07:1961-04:2014.

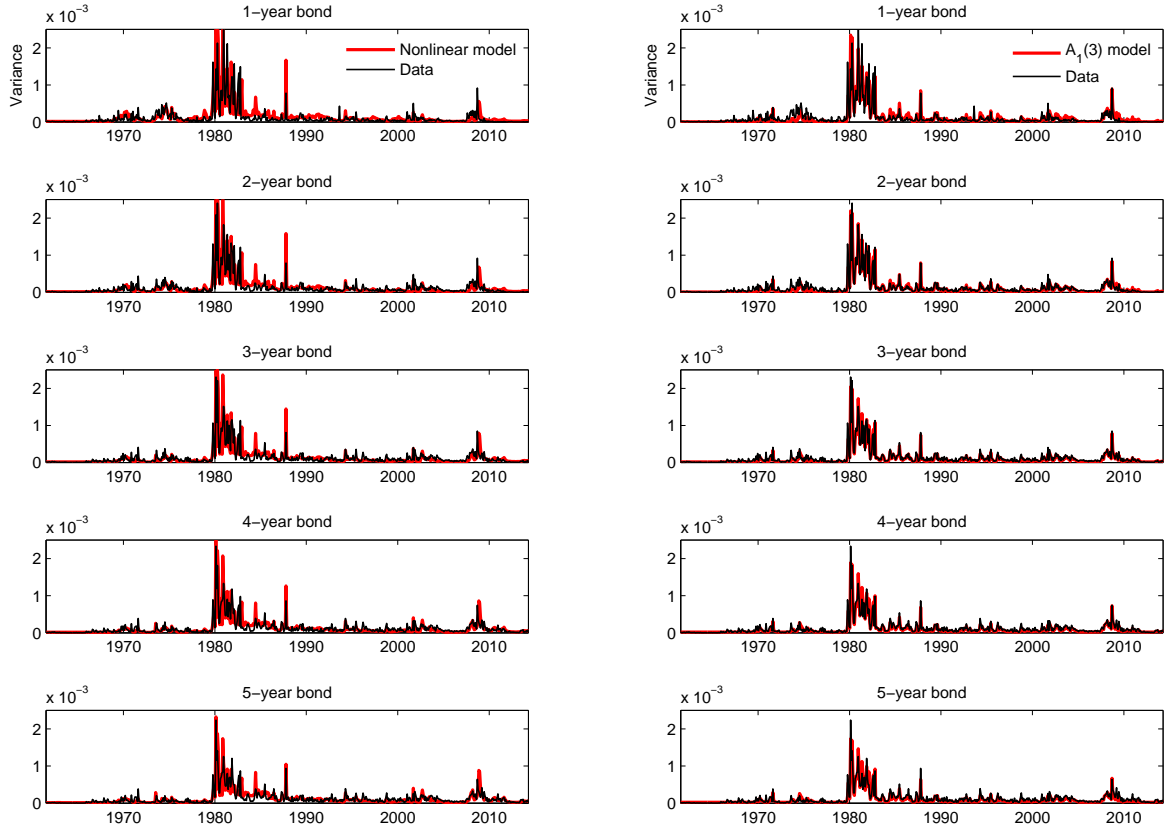


Figure 3: *Fit to realized variance.* This figure shows the realized annualized yield variance along with the (annualized) instantaneous yield variance in the nonlinear and the $A_1(3)$ model. The realized variance is calculated on a monthly basis and is based on daily squared yield changes. We use daily data from Gurkaynak, Sack, and Wright(2007) for the period 07:1961-04:2014.

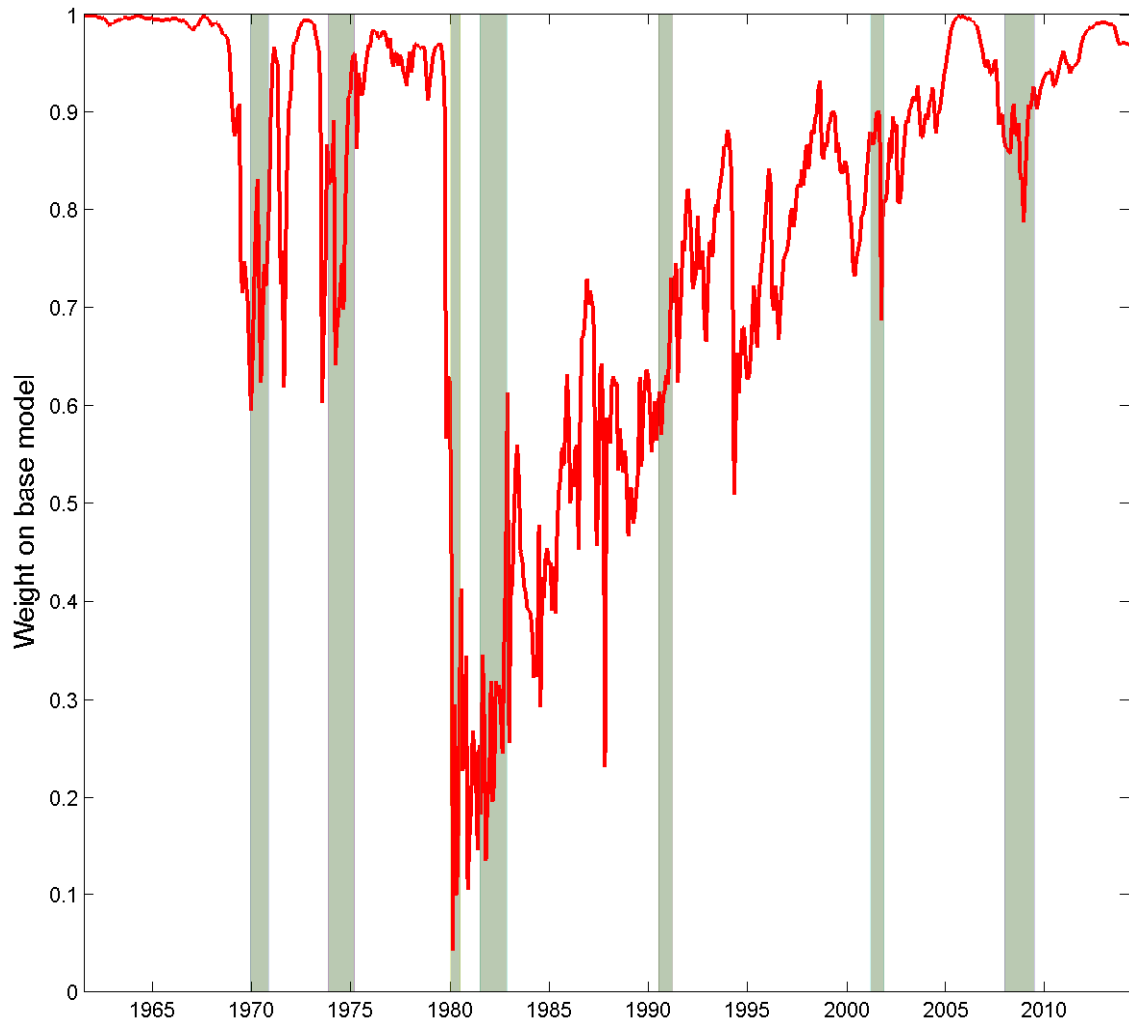


Figure 4: *Stochastic weight on Gaussian base model.* The bond price in the nonlinear model is $P(t, T) = s(t)P_0(t, T) + (1 - s(t))P_1(t, T)$ where $P_0(t, T)$ and $P_1(t, T)$ are bond prices that belong to the class of essentially affine Gaussian term structure models and $s(t)$ is a stochastic weight between 0 and 1. This figure shows the stochastic weight and the shaded areas show NBER recessions.

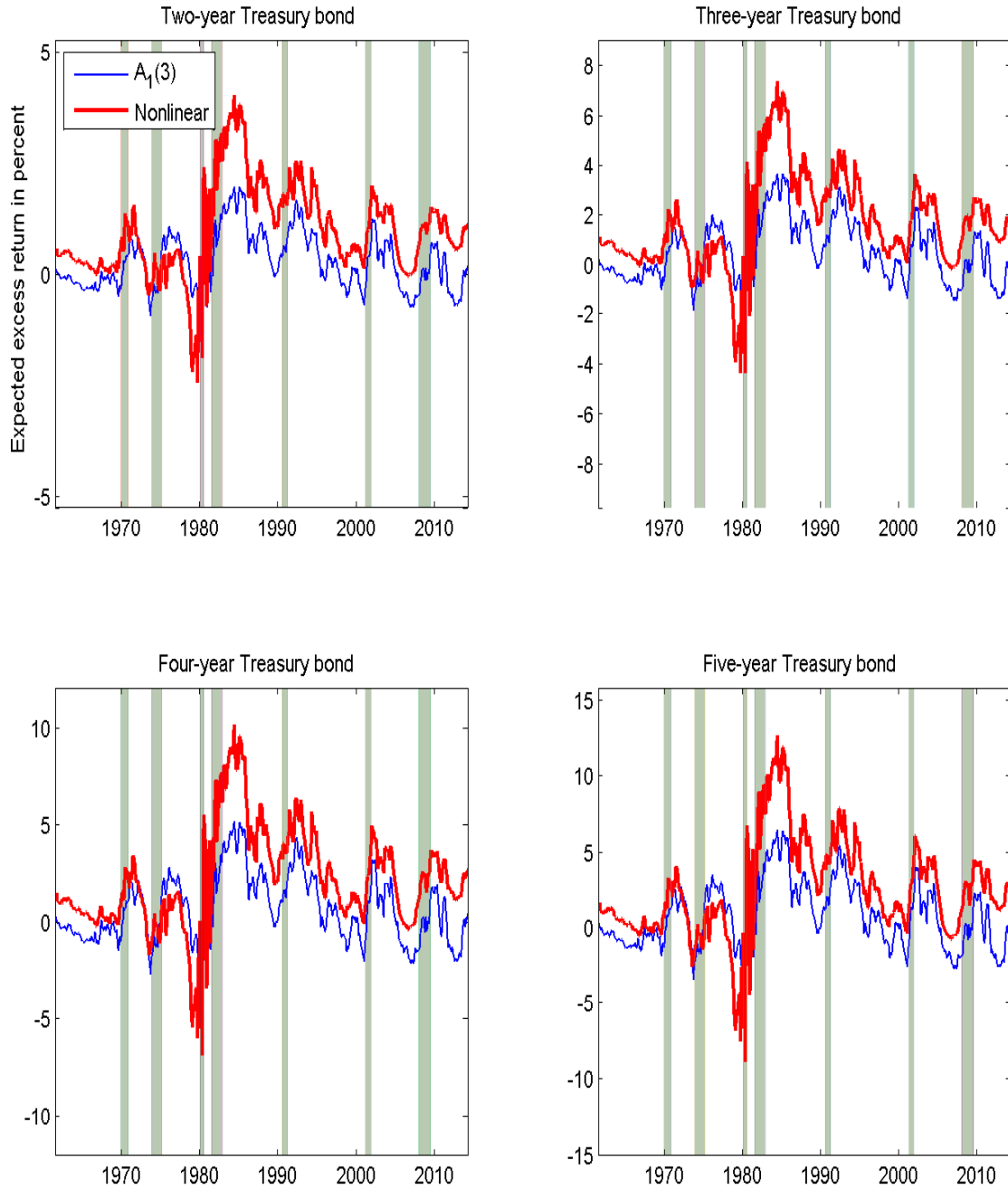


Figure 5: *Expected excess returns*. The graphs show the expected one year log excess returns of zero-coupon Treasury bonds with maturities of 2, 3, 4, and 5 years. The thin blue lines show expected excess returns in the three-factor $A_1(3)$ model and the thick red lines show expected excess returns in the three-factor nonlinear model. The shaded areas show NBER recessions.

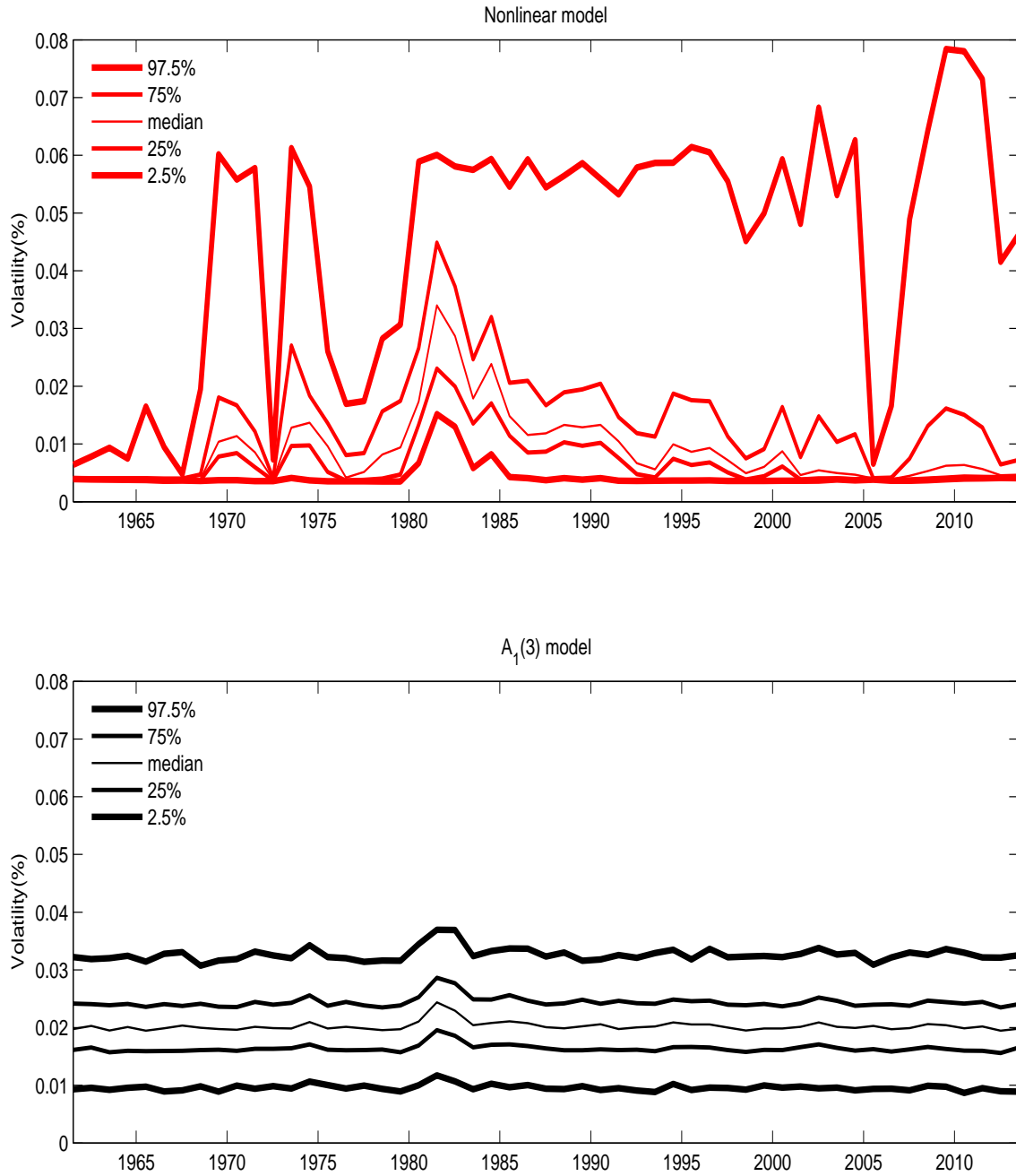


Figure 6: *Distribution of one-year ahead yield volatility.* The graphs show quantiles in the one-year ahead distribution of instantaneous volatility for the bond with a maturity of three years. The top graph shows the distribution in the three-factor nonlinear model, while the bottom graph shows the distribution in the three-factor $A_1(3)$ model. The data sample is 07:1961 to 04:2014 and the results for July in each year are plotted.

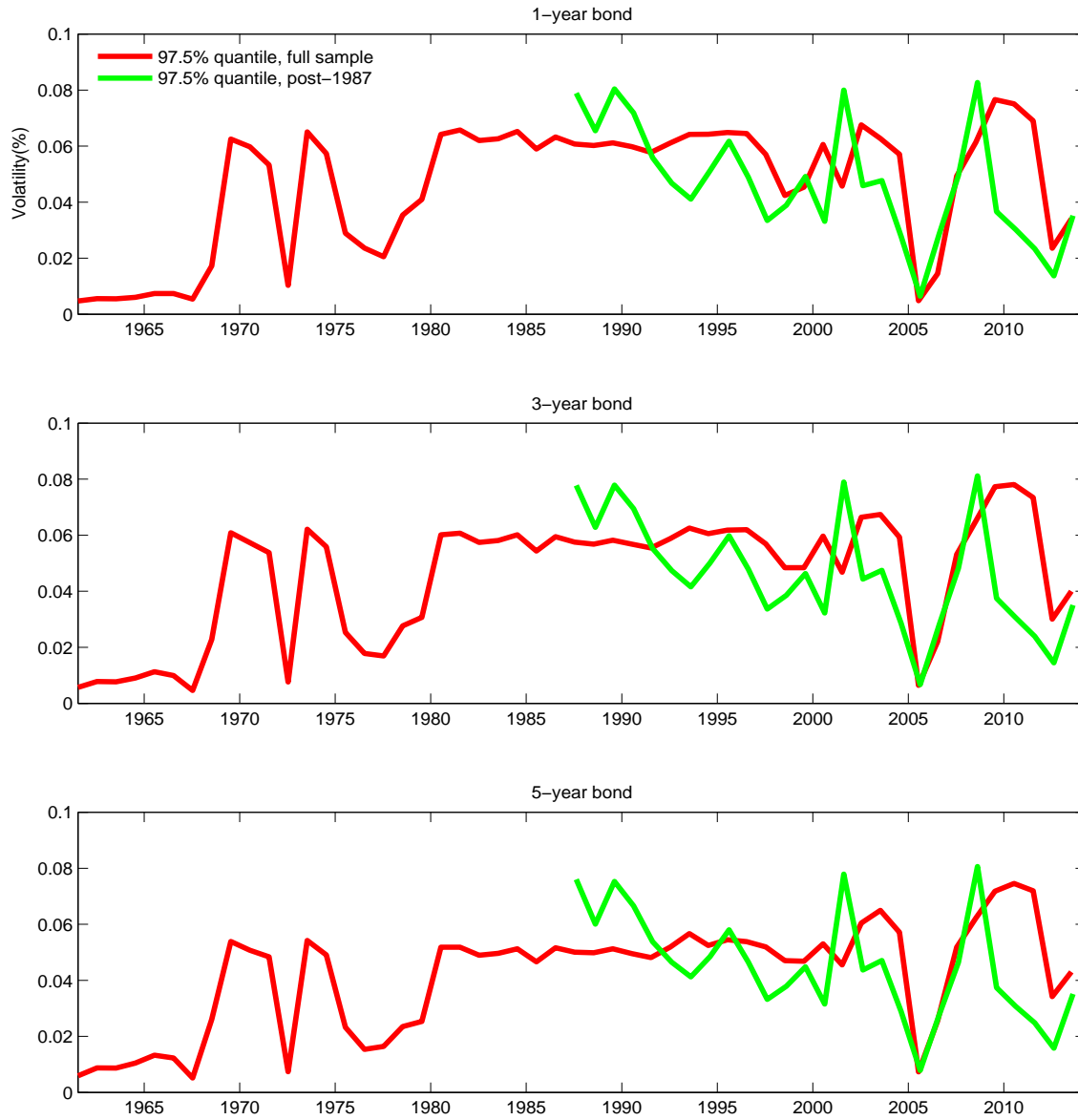


Figure 7: *Distribution of one-year ahead yield volatility for nonlinear model estimated using 1961-2014 and estimated using 1987-2014.* The graphs show the 97.5% quantiles in the one-year ahead distribution of instantaneous volatility. The red line shows the 97.5% quantiles in the three-factor nonlinear model, where the model is estimated by using data in the whole sample period 1961-2014. The yellow line shows the 97.5% quantiles in the three-factor nonlinear model, where the model is estimated by using data in the period 1987-2014. The results for September in each year are plotted.