

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

September 2016

Review of Finance, forthcoming

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).

Keywords: Nonlinear Term Structure Models, Expected Excess Returns, Stochastic Volatility, Unspanned Risk Premia (URP), Unspanned Stochastic Volatility (USV).

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; Glenn D. Rudebusch, 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 EFA 2015 In Vienna, WFA 2016 in Park City, Centre for Applied Macro-Finance and Asset Pricing Workshop (University of York), the Tripartite Conference held at Wharton, the World Finance Conference in Cyprus, the UBC Summer conference in Vancouver, the Inquire UK conference, and the Eurofdai 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

1. Introduction

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

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 et al. (2014) on unspanned expected excess returns, Jacobs and Karoui (2009) and Collin-Dufresne et al. (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 et al. (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 et al. (2009) and Creal and Wu (2015)). See also Dai and Singleton (2003) and Duffee (2010) and the references therein.

overfitting.

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 three-factor affine models with no or one stochastic volatility factor, the essentially affine $A_0(3)$ and $A_1(3)$ models.

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, 6.5% for the $A_1(3)$ model, 8% for the $A_0(3)$ 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 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 et al. (2014)

³See Ludvigson and Ng (2009), Cooper and Priestley (2009), Cieslak and Povala (2015), Duffee (2011b), Joslin et al. (2014), Chernov and Mueller (2012), and Bauer and Rudebusch (2016).

use five-factor Gaussian models where one or two factors that 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 era 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. The volatility in the Gaussian $A_0(3)$ model is constant and thus this model overestimates volatility during the Greenspan era and underestimates it during 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

⁴Papers on this topic include Collin-Dufresne and Goldstein (2002), Heidari and Wu (2003), Fan et al. (2003), Li and Zhao (2006), Carr et al. (2009), Andersen and Benzoni (2010), Bikbov and Chernov (2009), Joslin (2014), and Creal and Wu (2015).

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 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 et al. (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

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.

parameters from 23 to 27 whereas the regime-switching model in Dai et al. (2007) has 56 parameters. Quadratic term structure models have been proposed by Ahn et al. (2002) and Leippold and Wu (2003) among others, but Ahn et al. (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 et al. (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 et al. (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 et al. (2009) use the linearity generating framework of Gabaix (2009) to price swaps and interest rate derivatives. Similarly, in concurrent work Filipovic et al. (2015) introduce a linear-rational framework to price bonds and interest rate derivatives. Both approaches lead to closed form solutions of discount bonds, but their pricing framework is based on the potential approach of Rogers (1997) while our approach is based on a large class of equilibrium models discussed in Appendix B.⁶

The rest of the paper is organized as follows. Section 2 motivates and describes the model. Section 3 estimates the model and Section 4 presents the empirical results. In Section 5 we estimate a one-factor version of the nonlinear model and describe how nonlinearity works in this simple case, while Section 6 concludes.

⁶It is also possible to combine the general exponential-type stochastic discount factor (SDF) in our paper with the affine-type SDF in Filipovic et al. (2015) to get an exponential polynomial-type SDF similar to the setting of Chen and Joslin (2012).

2. A Nonlinear Term Structure Model

In this section we present a nonlinear model of the term structure of interest rates. We first motivate the model by presenting regression evidence for nonlinearities in excess returns and yield variances in Section 2.1 and then we present the model in Section 2.2.

2.1 MOTIVATING REGRESSION EVIDENCE

In Panel A of Table I we regress yearly excess returns measured on a monthly basis for the period 1961-2014 on the first three Principal Components (PCs) of yields and product combinations of the PCs. Specifically, the dependent variable is the average one-year excess return computed over US Treasury bonds with a maturity of 2, 3, 4, and 5 years (we explain the details of the data in Section 3.1). As independent variables, we first include all terms that are a product of up to three terms of the first three PCs (in short PC_1 , PC_2 , and PC_3). We then exclude terms with the lowest t -statistics one-by-one until only significant terms remain. The first row of Panel A shows the result. There are only three significant terms in the regression and they are all nonlinear. The second row shows the regression when we include only the first three PCs, the linear relation implied by affine models, and we see that the R^2 of 16% is substantially lower than the R^2 of 29% in the first regression. Finally, the third row shows that the linear terms add almost no explanatory power to the first regression.

Panel B in Table I shows similar regressions with the average excess return replaced by the average monthly realized yield variance as dependent variable (again, we leave the detailed explanation of how we calculate realized variance to Section 3.1). The first regression in Panel B shows the regression result when the independent variables are products of up to three terms of PC_1 , PC_2 , and PC_3 , after excluding

insignificant terms as in Panel A. None of the linear terms are significant and the five significant nonlinear terms generate an R^2 of 55%. Row two shows that a regression with only the first three PCs, the linear relation implied by affine models, yields a substantially lower R^2 of 34% and row three shows that the linear terms do not raise the R^2 when included in the first regression in Panel B.

These regressions show that there is a nonlinear relation both between yields and excess returns and between yields and yield variances. While the R^2 s in the nonlinear regressions are informative about the importance of nonlinearity, overfitting and collinearity limits the ability to pin down the precise nonlinear relation. In particular, when running the regressions for each bond maturity individually it is rare that the same set of nonlinear terms are significant. This evidence suggests that we need a parsimonious nonlinear model to study the nonlinearities in the first and second moments of bond returns, which we present in the next section.

2.2 THE MODEL

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

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

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

2.2.1. *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 et al. (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.⁷ 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 B 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 $d \times d$ -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.

⁷Constantinides (1992), Rogers (1997), Gabaix (2009), Carr et al. (2009), and Filipovic et al. (2015) also specify the functional form of the SDF directly and provide closed form solutions for bond prices.

2.2.2. 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,

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

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 for a more general class of nonlinear models and also show how our nonlinear model is related to the class of reduced form asset pricing model presented in Duffie et al. (2000) and Chen and Joslin (2012). 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.

2.2.3. *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.⁸

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 variations 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.

⁸Chan et al. (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.

2.2.4. *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 $P_0(t, T)$.⁹

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.¹⁰ Second, the short rate $r(t)$, the instantaneous volatility $v(t, T)$, and the instantaneous expected excess return $e(t, T)$

⁹If $\lambda_{0,X}$ and κ are zero, then $\sigma_0(T - t) = \sigma_1(T - t)$.

¹⁰The instantaneous volatility of the bond yield is $\frac{1}{T}v(t, t + \tau)$.

are nonlinear functions of $X(t)$.

3. Estimation

In this section, we estimate the nonlinear model described in Section 2 and compare it to standard essentially affine $A_0(3)$ and $A_1(3)$ models. All three models have three factors and the number of parameters is 22 in the $A_0(3)$ model, 23 in the $A_1(3)$ model and 26 in the nonlinear model. The $A_0(3)$ is a special case of our nonlinear model where $M_0(t) = M(t)$. The $A_1(3)$ model is well know and thus we only present the setup with results in Section 3.2 and defer details to Feldhütter (2016).

3.1 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. Although it is in theory sufficient to use bond yields to estimate the model, we add realized variances in the estimation to improve the identification of model parameters (see Cieslak and Povala (2016) for a similar approach). 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 et al. (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,

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

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

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 et al. (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.

3.2 THE $A_1(3)$ MODEL

We briefly describe the $A_1(3)$ model in this section and refer the reader to Feldhütter (2016) for a detailed discussion. The dynamics of the three-dimensional state vector $X(t) = (X_1(t), X_2(t), X_3(t))'$ are

$$dX(t) = \kappa (\bar{X} - X(t)) dt + S(t) dW(t), \quad (32)$$

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.

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

$$\kappa = \begin{pmatrix} \kappa_{(1,1)} & 0 & 0 \\ \kappa_{(2,1)} & \kappa_{(2,2)} & \kappa_{(2,3)} \\ \kappa_{(3,1)} & \kappa_{(3,2)} & \kappa_{(3,3)} \end{pmatrix} \quad (33)$$

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

$$S(t) = \begin{pmatrix} \sqrt{\delta_1 X_1(t)} & 0 & 0 \\ 0 & \sqrt{1 + \delta_2 X_1(t)} & 0 \\ 0 & 0 & \sqrt{1 + \delta_3 X_1(t)} \end{pmatrix} \quad (34)$$

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

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

$$\frac{dM(t)}{M(t)} = -r(t) dt - \Lambda(t)' dW(t), \quad (35)$$

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

$$r(t) = \rho_0 + \rho_X' X(t), \quad (36)$$

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

$$S(t)\Lambda(t) = \begin{pmatrix} \lambda_{X,(1,1)} X_1(t) \\ \lambda_{0,2} + \lambda_{X,(2,1)} X_1(t) + \lambda_{X,(2,2)} X_2(t) + \lambda_{X,(2,3)} X_3(t) \\ \lambda_{0,3} + \lambda_{X,(3,1)} X_1(t) + \lambda_{X,(3,2)} X_2(t) + \lambda_{X,(3,3)} X_3(t) \end{pmatrix}, \quad (37)$$

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

The bond price and the instantaneous yield volatility are

$$P(X(t), T) = e^{A(T-t) + B(T-t)'X(t)} \quad (38)$$

$$v(X(t), T) = \sqrt{B(T-t)'S(X(t))S(X(t))B(T-t)}, \quad (39)$$

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

$$\frac{dA(\tau)}{d\tau} = (\kappa\bar{X} - \lambda_0)' B(\tau) + \frac{1}{2} \sum_{i=2}^3 B_i(\tau)^2 - \rho_0, \quad A(0) = 0 \quad (40)$$

$$\frac{dB(\tau)}{d\tau} = (\kappa + \lambda_X)' B(\tau) + \frac{1}{2} \sum_{i=1}^3 B_i(\tau)\delta_i - \rho_X, \quad B(0) = 0_{3 \times 1}. \quad (41)$$

3.3 ESTIMATION METHODOLOGY

We use the Unscented Kalman Filter (UKF) to estimate the nonlinear model, the extended Kalman filter to estimate the $A_1(3)$ model, and the Kalman filter to estimate the $A_0(3)$ model. Christoffersen et al. (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 et al. (2014) and Carr and Wu (2009) for a detailed description of this nonlinear filter.

When we estimate the nonlinear and $A_1(3)$ model, 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 measurement equation is

$$\begin{pmatrix} Y_t \\ \text{RV}_t \end{pmatrix} = \begin{pmatrix} f(X_t) \\ g(X_t) \end{pmatrix} + \begin{pmatrix} \sigma_y I_5 & 0 \\ 0 & \sigma_{rv} I_5 \end{pmatrix} \epsilon_t, \quad \epsilon_t \sim N(0, I_{10}), \quad (42)$$

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

$$f_\tau(X_t) = -\frac{1}{\tau} \ln(P(X_t, t + \tau)) \quad (43)$$

$$g_\tau(X_t) = \frac{1}{\tau^2} v^2(X_t, t + \tau) \quad (44)$$

with $P(X_t, t + \tau)$ and $v(X_t, t + \tau)$ given in Equation (7) and (29), respectively. In the $A_0(3)$ model yield volatility is constant and we therefore only include yields (and not realized variances) in the estimation.

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

$$X_{t+1} = C + DX_t + \eta_{t+1}, \quad \eta_t \sim N(0, Q), \quad (45)$$

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 (45) 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 volatility $v(X_t, t + \tau)$ in equation (43) and (44) of the $A_1(3)$ model are given in equation (38) and (39) in Section 3.2. 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

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.

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 Section 3.2.

3.4 ESTIMATION RESULTS

Estimated parameters with asymptotic standard errors (in parenthesis) are reported in Table II and III. The second column of Table II shows parameter estimates based on the whole sample (1961:07-2014:04) that includes the period of the monetary experiments where the one-year bond yield and its volatility exceeded 15% and 5%, respectively. We re-estimate the nonlinear model using only yield and volatility data for the period 1987:08-2014:04, which excludes the high yield and yield volatility regime during the early eighties.¹³ The third column of Table II shows that the estimated parameters for this period are similar to the estimated parameters for the whole sample period. In particular, the nonlinear parameters β and γ have the same sign and are of similar magnitude. The parameter estimates for the $A_1(3)$ and the $A_0(3)$ model are reported in Table III.

The bond price in the nonlinear model is a weighted average of two Gaussian bond prices (see Theorem 1). Figure 2 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

¹³Alan Greenspan became chairman of the Fed on August 11, 1987.

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.

4. Empirical Results

In this section we show that the nonlinear three-factor model captures time variation in expected excess bond returns and yield volatility. Moreover, the nonlinearity leads to unspanned risk premia (URP) and unspanned stochastic volatility (USV), an empirical stylized fact, that affine models cannot capture without knife-edge restrictions and additional state variables that describe variations in expected excess returns and yield variances but not yields. While nonlinearities help explain time-variation in excess returns and yield variances, we show in Section 4.3 that the amount of nonlinearity in the cross-section is small and thus our model retains the linear relation of

US-Treasury yields across maturities.

4.1 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. Specifically, for all bond maturities $\tau = 2, 3, 4, 5$ we have

$$y_{t+1}^{\tau-1} - y_t^\tau = \text{const} + \phi^\tau \left(\frac{y_t^\tau - y_t^1}{\tau - 1} \right) + \text{residual}, \quad (46)$$

where y_t^τ is the (log) yield at time t of a zero-coupon bond maturing at time $t + \tau$. 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. Table IV replicates their findings for the sample period 1961:07-2014:04, that is, slope coefficients are negative, decreasing with maturity, and significantly different from one.

To check whether each model can match this stylized fact, we simulate a sample path of 1,000,000 months for two-, three-, four- and five-year excess bond returns and compare the model implied CS regression coefficients with those observed in the data. Table IV shows that the nonlinear model and $A_0(3)$ model captures the negative CS regression coefficients in population.

Figure 3 shows that one-year expected excess returns in the nonlinear model are negative in the early eighties and positive since the mid-80s while they are alternating between positive and negative in the $A_1(3)$ model. Expected excess returns in the $A_0(3)$ model are also positive since the mid-80s but both affine models cannot capture the very low and high realized excess returns during the monetary experiment. To

formally test whether the nonlinear model captures expected excess returns better than the two affine models we run regressions of realized excess returns on model implied expected excess returns in sample. Specifically,

$$rx_{t,t+n}^\tau = \alpha^{\tau,n} + \beta^{\tau,n} E_t [rx_{t,t+n}^\tau] + \text{residual}, \quad \forall \tau > n = 1, 2, 3, 4, 5, \quad (47)$$

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.¹⁴ The estimated expected excess returns for the nonlinear, $A_1(3)$, and $A_0(3)$ model are based on the sample period 1961:07 to 2014:04. The regression results are reported in Table V. If the model captures expected excess returns well, then the slope coefficient should be one, the constant zero. The slope coefficients are lower but generally close to one in the nonlinear model. In the $A_1(3)$ model the slope coefficients are close to one at the one-year horizon but are too low at longer horizon, while in the $A_0(3)$ model the slope coefficients are too high at the one-year horizon and too low for the three-and four year horizon. The average R^2 across bond maturity and holding horizon is 27.4% in the nonlinear model while it is only 6.5% in the $A_1(3)$ and 7.8% in the $A_0(3)$ 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 explanatory power of the model implied in sample expected excess return as follows¹⁵

$$\text{FVE} = 1 - \frac{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - E_t [rx_{t,t+n}^\tau])^2}{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^\tau - \frac{1}{T} \sum_{t=1}^T rx_{t,t+n}^\tau)^2}. \quad (48)$$

¹⁴Moments of yields and returns in the nonlinear model are easily calculated using Gauss-Hermite quadrature, see Appendix C 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.

¹⁵Almeida et al. (2011) refer to this measure as a modified R^2 .

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. Table V shows the FVEs of the nonlinear, $A_1(3)$, and $A_0(3)$ model for the sample period 1961:07 to 2014:04. The in sample FVEs for the nonlinear model are higher than for the $A_1(3)$ and $A_0(3)$ model. In contrast to the nonlinear and $A_0(3)$ 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, shown in the second to last column of Table V, is an upper bound for the FVE of any affine model for which expected excess returns are spanned by yields, e.g. the Cochrane and Piazzesi (2005) factor.¹⁶ The FVEs of the nonlinear model are equal to or higher than the explanatory power of the Cochrane-Piazzesi factor. 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. The last column of Table V shows that the explanatory power of any estimator for expected excess returns that is spanned by yields and their variances is lower than the FVE of our nonlinear model.

4.1.1. *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 (2016) find that the R^2 when regressing realized excess returns on the first three PC of yields along with expected inflation is 85%

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).

¹⁷See Ludvigson and Ng (2009), Cooper and Priestley (2009), Cieslak and Povala (2015), Duffee (2011b), Joslin et al. (2014), and Chernov and Mueller (2012). Bauer and Rudebusch (2016) argue that this evidence can be explained by measurement error.

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 et al. (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 PC, first and second PC, ..., and all five PCs of model-implied yields for the sample period 1961:07-2014:04. Specifically, for all bond maturities $\tau = 2, 3, 4, 5$ we run the in sample URP regressions

$$E_t [rx_{t,t+1}^\tau] = \alpha^{\tau,1:n} + \sum_{i=1}^n \beta^{\tau,1:n} PC_{i,t} + \varepsilon_t^{\tau,1:n}, \quad \forall n = 1, 2, 3, 4, 5, \quad (49)$$

where $PC_{i,t}$ denotes the i -th principal component of all five yields (ordered by decreasing contribution to the total variation in yields). The in sample R^2 s of these regressions are reported in Panels B, C, and D of Table VI. Panel C and D show that by construction the first three PCs explain all the variation in expected excess returns in the $A_1(3)$ and $A_0(3)$ model since expected excess returns are linear functions of yields in affine models. Panel B shows that the first three PCs explain on average 69.4% of the variation of expected excess returns in the nonlinear model. That is, almost one third of the variation of expected excess returns is due to a nonlinear relation between expected excess returns and yields in sample.

¹⁸The R^2 is 0.36 in the former and 0.195 in the latter, see Bauer and Rudebusch (2016)'s Table 3. Joslin et al. (2014) present similar evidence.

Empirically, realized excess returns are invariably used in lieu of expected excess returns as dependent variable. Hence, for all bond maturities $\tau = 2, 3, 4, 5$ we run the URP regressions

$$rx_{t,t+1}^\tau = \alpha^{\tau,1:n} + \sum_{i=1}^n \beta^{\tau,1:n} \text{PC}_{i,t} + \text{residual}, \quad \forall n = 1, 2, 3, 4, 5. \quad (50)$$

Panel A in VI shows R^2 from regressions of realized excess returns on PCs of model-implied yields in the data based on the sample period 1961:07-2014:04. To check whether each model can match the actual R^2 from the URP regression, we simulate a sample path of 1, 000, 000 months for two- to five-year excess bond returns and one- to five-year bond yields and compare the model implied URP regression R^2 s to those observed in the data. Panel E, F, and G show the population R^2 for the nonlinear, $A_1(3)$, and $A_0(3)$ model, respectively. In contrast to both affine models the population R^2 in the nonlinear model are largely in line with the actual R^2 observed in the data.

The final column in Panel E, F, and G shows the population R^2 when we replace the model implied PCs in URP regression (50) with the model implied expected excess return, that is,

$$rx_{t,t+1}^\tau = \alpha^\tau + \beta^\tau E_t [rx_{t,t+1}^\tau] + \text{residual}, \quad \forall \tau = 2, 3, 4, 5. \quad (51)$$

In the nonlinear model the average (over all bond maturities) population R^2 in regression (50) when $n = 3$ is 81% higher than in regression (51), that is, 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 (2016). 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 in sample residuals from regressing expected excess returns on PCs in the nonlinear model (Panel B in Table VI) and regress them on expected inflation. Specifically, for all bond maturities $\tau = 2, 3, 4, 5$ we run the regression

$$\varepsilon_t^{\tau,1:n} = \alpha^{\tau,n} + \beta^{\tau,n} \pi_t + \text{residual}, \quad \forall n = 3, 4, 5, \quad (52)$$

where $\varepsilon_t^{\tau,1:n}$ is the residual from URP regression (49) and π_t is an estimator for expected inflation that is based on the Michigan Survey of Consumers (MSC).¹⁹ Table VII shows the R^2 , slope coefficient, and 12-lag Newey-West corrected t-statistics of regression (52). Expected inflation explains about 11% of the variation in sample URP residuals based on the first three PCs and it is statistically significant at the 5% level. The R^2 s increase to slightly less than 20% when adding the fourth PC. Expected inflation remains statistically significant even when considering in sample URP residuals based all five PCs. Hence, 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.

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.

¹⁹Expected inflation is measured as the cross-sectional average of one-year ahead price growth forecasts of consumers surveyed by the University of Michigan. MSC is a survey conducted on monthly frequencies covering a large cross-section of consumers and Ang et al. (2007) show that it is a good unbiased predictor of inflation.

4.2 STOCHASTIC VOLATILITY

Table VIII 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 4 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 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 where a high-volatility scenario was unlikely. This finding suggest that there is information

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 $\frac{1}{\tau}v^{(\tau)}(t)$ with $v^{(\tau)}(t)$ given in equation (29).

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. Figure 5 shows the 97.5 quantiles of the one-year ahead distribution of yield volatility for sample periods with (1961:07-2014:04) and without (1987:08-2014:04) 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 et al. (2007), Bansal and Zhou (2002), and Bansal et al. (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 et al. (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.

4.2.1. *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. Specifically, for each bond maturity $\tau = 1, 2, 3, 4, 5$ and number $n = 1, 2, 3, 4, 5$ of PCs we run the following USV regression in the data

$$rv_t^\tau = \alpha^\tau + \sum_{i=1}^n \beta_i^\tau PC_{i,t} + \varepsilon_t^\tau, \quad (53)$$

where, as in the previous section, $PC^{i,t}$ denotes the i -th principal component of all five yields (ordered by decreasing contribution to the total variation in yields). The R^2 s of these USV regressions in the data are reported in Panel A of Table IX. 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 yield variance in the data.²²

To assess the ability of the nonlinear model to capture USV we regress model-implied instantaneous yield variance the PCs of model-implied yields:

$$v(t, t + \tau)^2 = \alpha^\tau + \sum_{i=1}^n \beta_i^{(\tau)} PC_t^i + \varepsilon_t^\tau, \quad (54)$$

where $v(t, t + \tau)$ is given in equation (29). Panel B shows that the average in sample R^2 from USV regressions (54) on the first three PCs ($n = 3$) is with 42.5% not substantially higher than in the data. In contrast, Panel D shows that in the $A_1(3)$

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.

model the in sample R^2 is 100% once the first three PCs are included in the USV regression (54). Hence, the presence of nonlinearities give rise to low R^2 's in USV regressions.

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

$$\sigma(t, T)' \sigma(t, T) = \sigma_{vol}(t, T)' \sigma_{vol}(t, T) + \sigma_{lev}(t, T)' \sigma_{lev}(t, T) + 2\sigma_{vol}(t, T)' \sigma_{lev}(t, T) \quad (55)$$

While the average (across maturities) in sample R^2 from regressing the yield variance on the first five PCs of model-implied yields is only 59.2% (see Panel B of Table IX), the average in sample R^2 from regressing each component in (55) 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 (55) 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).

The actual R^2 of USV regression (53) reported in Panel A of Table IX are not directly comparable to the in sample R^2 of USV regression (54) reported in Panel B for the nonlinear model and Panel C in the $A_1(3)$ model because realized variance based on daily data is a noisy proxy for yield variance. To check whether the nonlinear model can quantitatively capture USV in the data, we simulate 1,000,000 months of daily data (with 21 days in each month), compute the monthly realized variance and monthly average yield, and run the same URP regression as in the data, that is, regression (53). Panel C shows that the population R^2 's for the nonlinear model

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 (55) increases, the second part in the same equation tends to decrease.

are very similar to the in sample R^2 of Panel B where we use instantaneous variance instead of realized variance, that is, the average R^2 is 39.8% when including the first three PCs. Hence, our results are robust to taking into account that realized variance based on daily data is a noisy proxy for instantaneous variance. Panel E shows that the average population R^2 is 45.8% in the $A_1(3)$ model when regressing realized variance on the first three PCs of yields, which brings the population R^2 s much closer to R^2 s in the data. However, the populations R^2 s when using only one or two PCs in the $A_1(3)$ model are zero which is strongly at odds with the data.²⁴

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 IX confirms the factor structure in the data because the first PC of the residuals $\varepsilon_t^1, \dots, \varepsilon_t^5$ of the USV regression (53) explains 91.8% of the total variation in the USV residuals. Similarly, the first PC explains 98% of the variation in the residuals of USV regression (54) 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.

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

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.

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.

4.3 LINEARITY IN THE CROSS-SECTION OF YIELDS

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 three models (based on one million simulated observations) are shown in Table X.

Panel A of Table X 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 X shows that the population 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.

5. One Factor Model - an Illustration

In this section we estimate a one-factor nonlinear model to highlight the role of nonlinearity in a simple setting. Table XI shows the estimated parameters with

asymptotic standard errors (in parenthesis) based on the sample period 1961:07-2014:04. Panel A of Figure 6 shows the stochastic weight $s(X)$, defined in equation (8), over the sample period. The dynamics of $s(X)$ in the one-factor model are similar to the dynamics in the three-factor model – shown in Figure 2 – although $s(X)$ moves closer to zero in the three-factor model.

Panel B of Figure 6 shows bond yields as a function of the state variable X . The relation between yields and X is close to linear for low X 's, while for high X 's the rate of change picks up and yields increase more rapidly with X . The reason is that $s(X)$ starts to move away from one as X increases as seen in Panel C and moreover, the speed with which $s(X)$ moves away from one increases for high X s. Hence, for a given change in X , yields respond more for a high X , that corresponds to a high yield environment than for a low X , that corresponds to a low yield environment. Taken together, yield variances must be substantially higher for high yield environments than for low yield environments, which Panel D indeed shows. Moreover, the nonlinear relation between yields and their variances shown in Panel D leads to USV. Specifically, Panel B in Table XIII shows that the first PC of yields only explains between 52% and 63% of yield variance in sample. In contrast, in any affine one-factor stochastic volatility model the R^2 is 100%.

Panel E of Figure 6 shows the relation between yields and instantaneous expected excess returns. In a standard affine one-factor model the relation is linear, but we see that in the nonlinear model there is a U-shaped relation. This nonlinearity creates Unspanned Risk Premia in the model. Indeed, Panel A in Table XIII shows that the first PC of yields only explains between 13.0% and 19.9% of the variation in expected excess returns. Given the U-shaped relation between excess returns and yields it is not surprising that the level factor does not have more explanatory power but it provides a stark contrast to one-factor affine models where the first PC always explains 100%.

Finally, Figure 6's Panel F shows that the relation between the yields themselves is approximately linear. Thus, although there are significant nonlinear effects in the time series of excess returns and yield volatilities, there is an approximately linear relation between yields in the cross section which is consistent with the data.

6. 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. We explore this in Feldhütter et al. (2016).

A. General Nonlinear Gaussian Model

In this section we provide closed form solutions for a more general class of nonlinear term structure models, prove Theorem 1, and relate our results to the class of reduced form asset pricing models with closed form solutions discussed in Duffie et al. (2000) and Chen and Joslin (2012).

A.1 THE STOCHASTIC DISCOUNT FACTOR

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

$$M(t) = M_0(t) \left(1 + \gamma e^{-\beta' X(t)}\right)^\alpha, \quad (56)$$

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

A.2 CLOSED FORM BOND PRICES

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*

$$P(t, T) = \sum_{n=0}^{\alpha} s_n(t) P_n(t, T), \quad (57)$$

where

$$P_n(t, T) = e^{A_n(T-t) + B_n(T-t)'X(t)}, \quad (58)$$

$$s_n(t) = \frac{\binom{\alpha}{n} \gamma^n e^{-n\beta'X(t)}}{(1 + \gamma e^{-\beta'X(t)})^\alpha}. \quad (59)$$

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 (56) can be expanded as

$$M(t) = \sum_{n=0}^{\alpha} M_n(t), \quad (60)$$

where

$$M_n(t) = \binom{\alpha}{n} \gamma^n e^{-n\beta'X(t)} M_0(t). \quad (61)$$

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

$$\frac{dM_n(t)}{M_n(t)} = -r_n(t) dt - \Lambda_n(t)' dW(t), \quad (62)$$

where

$$\Lambda_n(t) = \Lambda_0(t) + n\Sigma'\beta \quad (63)$$

$$r_n(t) = r_0(t) + n\beta'\kappa(\bar{X} - X(t)) - \frac{n^2}{2}\beta'\Sigma\Sigma'\beta - n\beta'\Sigma\Lambda_0(t). \quad (64)$$

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 $M_n(t)$

²⁵Similar expansions of the stochastic discount factor appear in Yan (2008), Dumas et al. (2009), Bhamra and Uppal (2014), and Ehling et al. (2016).

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

$$P_n(t, T) = e^{A_n(T-t) + B_n(T-t)'X(t)}, \quad (65)$$

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

$$P(t, T) = \sum_{n=0}^{\alpha} s_n(t) P_n(t, T), \quad (66)$$

where $s_n(t)$ is given in equation (59). □

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

A.3 EXPECTED RETURN AND BOND VOLATILITY

Applying Ito's lemma to equation (56) leads to the dynamics of the stochastic discount factor:

$$\frac{dM(t)}{M(t)} = -r(t) dt - \Lambda(t)'dW(t), \quad (67)$$

where

$$\begin{aligned} r(t) &= r_0(t) + \alpha(1-s(t))\beta'\kappa(\bar{X} - X(t)) - \alpha(1-s(t))\beta'\Sigma\Lambda_0(t) \\ &\quad - \frac{\alpha}{2}(1-s(t))(\alpha(1-s(t)) + s(t))\beta'\Sigma\Sigma'\beta. \end{aligned} \quad (68)$$

and

$$\Lambda(t) = \Lambda_0(t) + \alpha(1-s(t))\Sigma'\beta. \quad (69)$$

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

$$\omega_n(t, T) = \frac{P_n(t, T)s_n(t)}{P(t, T)}. \quad (70)$$

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

$$\frac{dP(t, T)}{P(t, T)} = (r(t) + \Lambda(t)' \sigma(t, T)) dt + \sigma(t, T)' dW(t), \quad (71)$$

where

$$\sigma(t, T) = \Sigma' \left(\sum_{n=0}^{\alpha} \omega_n(t, T) B_n(T - t) + \beta \left(\sum_{n=0}^{\alpha} n \omega_n(t, T) - \alpha(1 - s(t)) \right) \right). \quad (72)$$

A.4 LINK TO REDUCED FORM ASSET PRICING MODELS

How is this model related to the large literature on reduced form asset pricing models with closed form solutions? At a first glance it does not seem to be related because the Gaussian state dynamics of $X(t)$ under the data generating or physical measure are no longer Gaussian under the risk neutral measure \mathbb{Q} . Specifically,

$$dX(t) = (\kappa \bar{X} - \kappa X(t) - \Sigma \Lambda(t)) dt + \Sigma dW^{\mathbb{Q}}(t), \quad (73)$$

where $\Lambda(t)$, given in equation (69), is a nonlinear function of $X(t)$ and

$$d\mathbb{Q} = e^{-\frac{1}{2} \int_0^t \Lambda(a)' \Lambda(a) da - \int_0^t \Lambda(a)' dW^{\mathbb{P}}(t)} d\mathbb{P}. \quad (74)$$

However, we can compute the state dynamics under the risk neutral measure in the benchmark model defined as

$$d\mathbb{Q}^0 = e^{-\frac{1}{2} \int_0^t \Lambda_0(a)' \Lambda_0(a) da - \int_0^t \Lambda_0(a)' dW^{\mathbb{P}}(t)} d\mathbb{P}, \quad (75)$$

where $\Lambda_0(t)$, which is given in equation (19), is an affine function of $X(t)$ and thus Gaussian under \mathbb{Q}^0 . Specifically,

$$dX(t) = (\kappa \bar{X} - \Sigma \lambda_{0,0} - (\kappa + \Sigma \lambda_{0,X}) X(t)) dt + \Sigma dW^{\mathbb{Q}^0}(t). \quad (76)$$

Define

$$f(X_T) = \frac{(1 + \gamma e^{-\beta' X(T)})^\alpha}{(1 + \gamma e^{-\beta' X(t)})^\alpha} \quad (77)$$

and rewrite the bond price as an expectation under the risk neutral measure in the benchmark model. Specifically,

$$\begin{aligned} P(t, T) &= \mathbb{E}_t \left[\frac{M(T)}{M(t)} \right] = \mathbb{E}_t \left[\frac{M_0(T)}{M_0(t)} f(X_T) \right] \\ &= \mathbb{E}_t^{\mathbb{Q}^0} \left[e^{-\int_t^T r_0(a) da} f(X(T)) \right], \end{aligned} \quad (78)$$

where $r_0(t)$, given in equation (18), is affine in $X(t)$. Duffie et al. (2000) and Chen and Joslin (2012) show that the expectation in equation (78) can be solved in closed form if $f(x) = \sum_n (c_n + v_n x) e^{\beta_n x}$, the short rate is affine in $X(t)$, and $X(t)$ is Gaussian under \mathbb{Q} . As shown in the proof of Theorem 1, the function $f(X_t)$ can be expanded into the exponential polynomial

$$f(X_T) = \sum_{n=0}^{\alpha} \frac{\binom{\alpha}{n} \gamma^n}{(1 + \gamma e^{-\beta' X(t)})^\alpha} e^{-n\beta' X(T)} = \sum_{n=0}^{\alpha} v_n e^{-n\beta' X(T)}. \quad (79)$$

using the Binomial expansion theorem and hence the bond price is given in closed form.

B. Equilibrium Models

In this section we show that the functional form of the state price density in equation (2) and (56) 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

$$dX(t) = (\theta - \kappa X(t)) dt + \Sigma dW(t), \quad (80)$$

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

$$M_0(t) = e^{-\rho t} C(t)^{-R}, \quad (81)$$

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

$$\frac{dC(t)}{C(t)} = \mu_C dt + \sigma'_C dW(t). \quad (82)$$

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.

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

$$\Lambda_0 = R\sigma_C \quad (83)$$

$$r_0 = \rho + R\mu_C - \frac{1}{2}R(R+1)\sigma'_C\sigma_C. \quad (84)$$

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

B.1 TWO TREES

Cochrane et al. (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

$$dD_i(t) = D_i(t) (\mu_i dt + \sigma'_i dW(t)). \quad (85)$$

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

$$\begin{aligned} M(t) &= e^{-\rho t} C(t)^{-R} \\ &= e^{-\rho t} (D_1(t) + D_2(t))^{-R} \\ &= e^{-\rho t} D_1(t)^{-R} \left(1 + \frac{D_2(t)}{D_1(t)} \right)^{-R} \\ &= M_0(t) \left(1 + e^{\log(D_2(t)) - \log(D_1(t))} \right)^{-R}, \end{aligned} \quad (86)$$

where $M_0(t) = e^{-\rho t} D_1(t)^{-R}$ and $X(t) = \log(D_1(t)/D_2(t))$. Equation (86) has the same form as the SDF in equation (56) 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.

B.2 MULTIPLE CONSUMPTION GOODS

Models with multiple consumption goods and CES consumption aggregator naturally falls within the functional form of the SDF in equation (56). Consider a setting with two consumption goods. The aggregate output of the two goods are given by

$$dD_i(t) = D_i(t) (\mu_i dt + \sigma'_i dW(t)). \quad (87)$$

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

$$u(C, t) = e^{-\rho t} \frac{1}{1-R} C^{1-R}, \quad (88)$$

where

$$C(C_1, C_2) = \left(\phi^{1-b} C_1^b + (1-\phi)^{1-b} C_2^b \right)^{\frac{1}{b}}. \quad (89)$$

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

$$\begin{aligned} M(t) &= e^{-\rho t} C(t)^{-R} \\ &= (\phi)^{\frac{bR}{1-b}} e^{-\rho t} D_1(t)^{-R} \left(1 + \left(\frac{1-\phi}{\phi} \right)^{1-b} \left(\frac{D_2(t)}{D_1(t)} \right)^b \right)^{-\frac{R}{b}}. \end{aligned} \quad (90)$$

After normalizing equation (90) has the same form as the SDF in equation (56) 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.

B.3 EXTERNAL HABIT FORMATION

The utility function in Campbell and Cochrane (1999) is

$$U(C, H) = e^{-\rho t} \frac{1}{1-R} (C - H)^{1-R}, \quad (91)$$

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

$$M(t) = e^{-\rho t} C(t)^{-R} s(t)^{-R} \quad (92)$$

$$= M_0(t) s(t)^{-R}. \quad (93)$$

Define the state variable

$$dX(t) = \kappa (\bar{X} - X(t)) dt + b dW(t), \quad (94)$$

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

$$ds(t) = s(t) (\mu_s(t) dt + \sigma_s(t) dW(t)), \quad (95)$$

where

$$\mu_s(t) = (1 - s(t)) \left(\beta \kappa (\bar{X} - X(t)) + \frac{1}{2} (1 - 2s(t)) \beta^2 b^2 \right) \quad (96)$$

$$\sigma_s(t) = (1 - s(t)) \beta b. \quad (97)$$

The functional form of the surplus consumption ratio differs from Campbell and Cochrane (1999). However, note that the surplus consumption ratio is locally per-

fectly 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

$$M(t) = M_0(t) \left(1 + e^{-\beta X(t)}\right)^R. \quad (98)$$

The above state price density has the same form as equation (56) 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.

B.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 (82). 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),

$$U(C, \lambda) = \max_{\{C_1+C_2=C\}} \left(\frac{1}{1-R} C_1^{1-R} + \lambda \frac{1}{1-R} C_2^{1-R} \right). \quad (99)$$

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

$$C_1(t) = s(t)C(t), \quad (100)$$

$$C_2(t) = (1 - s(t)) C(t), \quad (101)$$

²⁷The model can easily be generalised to a setting with disagreement about multiple stochastic processes and learning. For instance, Ehling et al. (2016) show that in a model with disagreement about inflation, the bond prices are weighted averages of quadratic Gaussian term structure models.

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

$$\begin{aligned}
M(t) &= e^{-\rho t} C_1(t)^{-R} \\
&= e^{-\rho t} C(t)^{-R} s(t)^{-R} \\
&= M_0(t) \left(1 + e^{\frac{1}{R} \log(\lambda(t))}\right)^R.
\end{aligned} \tag{102}$$

This has the same form as equation (56) 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.

B.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

$$\begin{aligned}
M(t) &= e^{-\rho t} C(t)^{-R} \\
&= e^{-\rho t} (C(t) + b)^{-R} \\
&= e^{-\rho t} C(t)^{-R} \left(1 + \frac{b}{C(t)}\right)^{-R} \\
&= M_0(t) \left(1 + e^{\log(b) - \log(C(t))}\right)^{-R}
\end{aligned} \tag{103}$$

After normalizing equation (103) has the same form as the SDF in equation (56) 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.

C. 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 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

$$E_t[f(X(u))] = \int_{\mathcal{R}^d} f(x) \frac{1}{\left((2\pi)^d |\Sigma_X|\right)^{0.5}} e^{-\frac{1}{2}(x-\mu_X)'\Sigma_X^{-1}(x-\mu_X)} dx. \quad (104)$$

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 (104) as

$$\pi^{-\frac{d}{2}} \int_{\mathcal{R}^d} f(\sqrt{2}\sigma_X y + \mu_X) e^{-y'y} dy. \quad (105)$$

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 (105) can be approximated by the n point Gauss-Hermite quadrature

$$\int_{\mathcal{R}^d} f(\sqrt{2}\sigma_X y + \mu_X) e^{-y'y} dy \approx \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_i w_j w_k g(y_1(i), y_2(j), y_3(k)), \quad (106)$$

where w_i are the weights and $y_l(i)$ are the nodes for the n point Gauss-Hermite quadrature.

²⁸For more details see Judd (1998).

ture for $i = 1, \dots, n$ and $l = 1, \dots, 3$. We use $n = 4$ in equation (106).

References

- Ahn, D.-H., R. Dittmar, and A. Gallant, 2002, Quadratic term structure models: Theory and evidence, *Review of Financial Studies* 15, 243–288.
- Ahn, D.-H., R. F. Dittmar, A. R. Gallant, and B. Gao, 2003, Purebred or hybrid? Reproducing the volatility in term structure dynamics, *Journal of Econometrics* 116, 147–180.
- Ahn, D.-H., and B. Gao, 1999, A parametric non-linear model of term structure dynamics, *Review of Financial Studies* 12, 721–762.
- Ait-Sahalia, Y., 1996a, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527–560.
- Ait-Sahalia, Y., 1996b, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.
- Almeida, C., J. J. Graveline, and S. Joslin, 2011, Do interest rate options contain information about excess bond returns?, *Journal of Econometrics* 164, 35–44.
- Andersen, T. G., and L. 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, T. G., T. Bollerslev, and F. X. Diebold, 2010, Parametric and nonparametric volatility measurement, in Y. Ait-Sahalia, and L. P. Hansen, eds., *Handbook of Financial Econometrics, Vol 1: Tools and Techniques (Handbooks in Finance)*, 67–137 (Elsevier).

- Ang, A., and G. Bekaert, 2002, Short rate nonlinearities and regime switches, *Journal of Economic Dynamics and Control* 26, 1243–1274.
- Ang, A., G. Bekaert, and M. Wei, 2007, Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* 54, 1163–1212.
- Balduzzi, P., and I.-H. E. Chiang, 2012, A simple test of the affine class of term structure models, *Review of Asset Pricing Studies* 2, 203–244.
- Bansal, R., G. Tauchen, and H. Zhou, 2004, Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle, *Journal of Business and Economic Statistics* 22, 396–409.
- Bansal, R., and H. Zhou, 2002, Term structure of interest rates with regime shifts, *Journal of Finance* 57, 1997–2043.
- Barndorff-Nielsen, O., and N. 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, S., 2000, A model of dynamic equilibrium asset pricing with heterogeneous beliefs and extraneous risk, *Journal of Economic Dynamics and Control* 24, 63–95.
- Bauer, M., and G. Rudebusch, 2016, Resolving the spanning puzzle in macro-finance term structure models, *forthcoming, Review of Finance* .
- Bhamra, H., and R. 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* 1292–1305.

- Campbell, J. Y., and J. 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, J. Y., and R. J. Shiller, 1991, Yield spread and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495–514.
- Carr, P., X. Gabaix, and L. Wu, 2009, Linearity-generating processes, unspanned stochastic volatility, and interest-rate option pricing, Working paper.
- Carr, P., and L. Wu, 2009, Stock options and credit default swaps: A joint framework for valuation and estimation, *Journal of Financial Econometrics* 1–41.
- Chan, K., A. Karolyi, F. Longstaff, and A. B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.
- Chapman, D. A., and N. D. Pearson, 2000, Is the short rate drift actually nonlinear?, *Journal of Finance* LV, 355–388.
- Chen, H., and S. Joslin, 2012, Generalized transform analysis of affine processes and applications in finance, *Review of Financial Studies* 25, 2225–2256.
- Chernov, M., and P. Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.
- Christoffersen, P., C. Dorion, K. Jacobs, and L. Karoui, 2014, Nonlinear Kalman Filtering in Affine Term Structure Models, *Management Science* 60.
- Cieslak, A., and P. Povala, 2015, Expected returns in treasury bonds, conditionally accepted, *Review of Financial Studies*.
- Cieslak, A., and P. Povala, 2016, Information in the term structure of yield curve volatility, *Journal of Finance* 71, 1393–1436.

- Cochrane, J., and M. 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, P., and R. S. Goldstein, 2002, Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility, *Journal of Finance* 57, 1685–1730.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2009, Can interest rate volatility be extracted from the cross section of bond yields?, *Journal of Financial Economics* 94, 47–66.
- Constantinides, G. M., 1992, A theory of the nominal term structure of interest rates, *Review of Financial Studies* 5, 531–552.
- Cooper, I., and R. Priestley, 2009, Time-Varying Risk Premiums and the Output Gap, *Review of Financial Studies* 22, 2601–2633.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, p. 385–408.
- Creal, D. D., and J. C. Wu, 2015, Estimation of affine term structure models with spanned or unspanned stochastic volatility, *Journal of Econometrics* 185, 60–81.
- Dai, Q., A. Le, and K. 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.
- Dai, Q., and K. Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631–678.

- Dai, Q., and K. J. Singleton, 2002, Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of Financial Economics* 63, 415–441.
- Dai, Q., K. J. Singleton, and W. Yang, 2007, Regime shifts in a dynamic term structure model of u.s. treasury bond yields, *Review of Financial Studies* 20, 1669–1706.
- Duffee, G., 2002, Term premia and interest rate forecast in affine models, *Journal of Finance* 57, 405–443.
- Duffee, G., 2010, Sharpe ratios in term structure models, *Working Paper, University of California, Berkeley* .
- Duffee, G., 2011a, Forecasting with the term structure: The role of no-arbitrage restrictions, *Working Paper, Johns Hopkins University* .
- Duffee, G., 2011b, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895–2934.
- Duffee, G., 2012, Forecasting interest rates, Working paper.
- Duffie, D., and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, p. 379–406.
- Duffie, D., J. Pan, and K. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.
- Dumas, B., A. Kurshev, and R. Uppal, 2009, Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility, *The Journal of Finance* 64, 579–629.
- Ehling, P., M. Gallmeyer, C. Heyerdahl-Larsen, and P. Illeditsch, 2016, Disagreement about inflation and the yield curve, Working Paper.

- Fama, E. F., and R. R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.
- Fan, R., A. Gupta, and P. Ritchken, 2003, Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets, *Journal of Finance* 58, 2219–2248.
- Feldhütter, P., 2016, Can affine models match the moments in bond yields?, *Quarterly Journal of Finance* 6, 1–56.
- Feldhütter, P., C. Heyerdahl-Larsen, and P. Illeditsch, 2016, Expanded term structure models, Work in progress.
- Feldhütter, P., and D. Lando, 2008, Decomposing Swap Spreads, *Journal of Financial Economics* 88, 375–405.
- Filipovic, D., M. Larsson, and A. B. Trolle, 2015, Linear-Rational Term Structure Models, *forthcoming, Journal of Finance* .
- Gabaix, X., 2009, Linearity-generating processes: A modelling tool yielding closed forms for asset prices, CEPR and NBER Working paper.
- Gurkaynak, R. S., B. Sack, and J. H. Wright, 2007, The u.s. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291–2304.
- Hansen, L. P., and R. J. Hodrick, 1980, Forward Exchange Rates as Optimal predictors of future Spot Rates: An Econometric Analysis, *Journal of Political Economy* 88, 829–853.
- Heidari, M., and L. Wu, 2003, Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?, *Journal of Fixed Income* 12, 75–86.

- Jacobs, K., and L. Karoui, 2009, Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets, *Journal of Financial Economics* 91, 288–318.
- Jermann, U., 2013, A production-based model for the term structure, *Journal of Financial Economics* 109, 293–306.
- Jones, C. S., 2003, Nonlinear mean reversion in the short-term interest rate, *Review of Financial Studies* 16, 793–843.
- Joslin, S., 2014, Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?, *forthcoming, Management Science* .
- Joslin, S., M. Pribsch, and K. Singleton, 2014, Risk premiums in dynamic term structure models with unspanned macro risks, *JF* 69, 1197–1233.
- Judd, K. L., 1998, *Numerical Methods in Economics*, first edition (MIT Press).
- Leippold, M., and L. Wu, 2003, Design and Estimation of Quadratic Term Structure Models, *European Finance Review* 7, 47–73.
- Li, H., and F. Zhao, 2006, Unspanned stochastic volatility: Evidence from hedging interest rate derivatives, *Journal of Finance* 61, 341–378.
- Ludvigson, S., and S. Ng, 2009, Macro Factors in Bond Risk Premia, *Review of Financial Studies* 22, 5027–5067.
- Newey, W. K., and K. D. West, 1987, A simple, positive sem-definite, heteroscedasticity, and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Pritsker, M., 1998, Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies* 11, 449–487.
- Richard, S., 2013, A non-linear macroeconomic term structure model, Working Paper.

- Rogers, L., 1997, The potential approach to the term structure of interest rates and foreign exchange rates, *Mathematical Finance* 7, 157–164.
- Stanton, R., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* VII, 1973–2002.
- Tang, H., and Y. 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, J., 2011, Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset, *American Economic Review* 101, 1514–1534.
- Yan, H., 2008, Natural selection in financial markets: Does it work?, *Management Science* 54, 1935–1950.

Table I: *Nonlinearities in expected excess returns and realized variances.* This table shows coefficients, standard errors (in brackets), and R^2 s from regressions of realized one-year log excess bond returns (Panel A) and realized yield variances (Panel B), averaged over bond maturities two to five in Panel A and one to five in Panel B, on three different sets of yield principal components (PCs) and powers thereof. The independent variables in the first row of both panels are obtained by first considering all product combinations of the first three PCs up to and including order three and excluding every variable with the lowest t-statistic until only significant variables remain. The monthly excess returns, realized variances, and PCs are calculated using daily zero-coupon bond yield data from 1961:07 to 2014:04. The bond maturities are ranging from 1-5 years and the data are obtained from Gurkaynak, Sack, and Wright (2007). The number of observations is 622 for the predictive regressions in Panel A and 634 for the contemporaneous regressions in Panel B. All variables are standardized and standard errors are computed using the Hansen and Hodrick correction with 12 lags in Panel A and the Newey and West correction with 12 lags in Panel B. ** and * indicate statistical significance at the 1% and 5% levels, respectively.

Panel A: One-year average excess bond returns

PC_1	PC_2	PC_3	PC_1PC_2	PC_1^3	PC_2^3	R^2
			-0.37** (0.09)	0.40** (0.11)	0.33** (0.08)	0.29
0.07 (0.13)	0.39** (0.12)	-0.05 (0.11)				0.16
-0.14 (0.17)	0.10 (0.14)	-0.04 (0.10)	-0.33** (0.10)	0.49** (0.16)	0.26* (0.11)	0.30

Panel B: Realized average yield variance

PC_1	PC_2	PC_3	PC_1^2	PC_1PC_3	PC_2PC_3	PC_1^3	$PC_1PC_2PC_3$	R^2
			0.12** (0.04)	-0.12* (0.05)	-0.18** (0.06)	0.39** (0.07)	-0.34** (0.05)	0.55
0.48** (0.12)	-0.10 (0.09)	0.32** (0.09)						0.34
0.10 (0.14)	0.04 (0.05)	0.04 (0.06)	0.14 (0.08)	-0.10 (0.05)	-0.16* (0.07)	0.30* (0.15)	-0.34** (0.08)	0.55

Table II: *Parameter estimates of the nonlinear three-factor model.* This table contains parameter estimates and asymptotic standard errors (in parenthesis) for the nonlinear three-factor model. The left column shows parameters estimates based on yield and realized variance data for the whole sample (1961:07-2014:04) and the right column shows parameter estimates based on yield and realized variance data for the Post-Volcker period (1987:08-2014:04). The bond maturities are ranging from 1-5 years and the data are obtained from Gurkaynak, Sack, and Wright (2007). The unscented Kalman filter is used to estimate the nonlinear model.

	Nonlinear Model (1961-2014)			Nonlinear Model (1987-2014)		
	0.3127 (0.04224)	0	0	0.3452 (0.08753)	0	0
κ	0.3063 (0.05601)	0.002189 (2.246e-05)	0	0.5507 (0.09825)	0.003245 (0.002091)	0
	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)
ρ_0		-0.001756 (0.01408)			-0.001002 (0.02238)	
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)
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)
	-0.2187 (0.04129)	0.005572 (0.001321)	-0.02053 (0.005609)	-0.2244 (0.06907)	0.003604 (0.00792)	-0.02491 (0.04552)
λ_X	-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)
	-0.2943 (0.1053)	-0.02387 (0.01562)	0.04613 (0.05121)	-0.3973 (0.2578)	-0.0237 (0.02542)	0.05947 (0.2159)
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)
σ_y		0.0005463 (6.945e-05)			0.0004679 (9.47e-05)	
σ_{rv}		7.281e - 05 (8.491e-06)			2.857e - 05 (3.381e-06)	

Table III: *Parameter estimates of the $A_1(3)$ and the $A_0(3)$ model.* This table contains parameter estimates and asymptotic standard errors (in parenthesis) for two three-factor affine models: the $A_1(3)$ model with one stochastic volatility factor and the $A_0(3)$ model with only Gaussian factors. The parameter estimates for the $A_1(3)$ model are based on yield and realized variance data for the whole sample (1961:07-2014:04) and the parameter estimates for the $A_0(3)$ model are based on yield data for the whole sample. The bond maturities are ranging from 1-5 years and the data are obtained from Gurkaynak, Sack, and Wright (2007). The extended Kalman filter is used to estimate the $A_1(3)$ model and the Kalman filter is used to estimate the $A_0(3)$ model.

	$A_1(3)$ Model (1961-2014)			$A_0(3)$ Model (1961-2014)		
	1.421 (0.1863)	0	0	0.7064 (0.1982)	0	0
κ	-0.04787 (1.899)	0.07225 (0.01938)	-0.003283 (4.101)	0.3558 (0.2189)	0.06629 (0.06185)	0
	0.283 (0.6523)	-0.009014 (0.07474)	0.356 (0.01893)	0.6473 (0.1987)	0.3549 (0.2011)	0.8202 (0.1865)
ρ_0		0.08832 (0.3038)			0.02046 (0.06848)	
ρ_X	0.0003736 (0.0002645)	0.001131 (0.0009603)	$1.385e - 05$ (0.000302)	-0.001232 (0.002566)	0.01626 (0.002255)	0.01085 (0.003361)
λ_0	0	0.6101 (106.4)	0.006454 (7.178)	0.1353 (0.1707)	-0.3741 (0.1998)	0.1233 (0.4018)
	$6.75e - 05$ (0.07544)	0	0	-0.335 (0.1954)	-0.01799 (0.03515)	0.006627 (0.09816)
λ_X	2.378 (3.64)	-0.0006549 (0.01964)	3.381 (5.878)	$-7.847e - 05$ (0.001684)	0.1821 (0.1682)	0.5751 (0.114)
	0.01683 (0.7003)	-0.0001671 (0.0733)	$1.302e - 05$ (0.01966)	0.183 (0.2063)	-0.09196 (0.08949)	-0.03485 (0.1974)
δ	0	491.5 (836.6)	2.417 (0.3336)			
$(\kappa\bar{X})$	1.509 (0.1109)	0	0			
σ_y		0.0006001 ($8.676e-05$)			0.0001038 ($1.698e-05$)	
σ_{rv}		$6.18e - 05$ ($6.019e-06$)				

Table IV: *Campbell-Shiller regressions*. This table shows the coefficients ϕ^τ from the regressions $y_{t+\tau}^{\tau-1} - y_t^\tau = \text{const} + \phi^\tau \left(\frac{y_t^\tau - y_t^1}{\tau-1} \right) + \text{residual}$, where y_t^τ 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). Standard errors in parentheses are computed using the Hansen and Hodrick correction with 12 lags. The population coefficients for each model are based on one simulated sample path of 1,000,000 months.

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	
$A_0(3)$ model	-0.18	-0.37	-0.54	-0.71	

Table V: *Excess return regressions.* This table shows regression coefficients from a regression of realized (log) excess returns on model implied expected (log) excess returns in sample. Monthly data for bonds with maturities ranging from one through five-years are from Gurkaynak, Sack, and Wright (2007) for the period 1961:07 to 2014:04. For $n = 1, 2, 3, 4, n$ -year excess returns are computed by subtracting the n -year (log) yield from the n -year (log) holding-period return of a τ -year bond ($\tau > n$). The fraction of variance explained is defined as $FVE = 1 - \frac{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^{\tau} - E_t[rx_{t,t+n}^{\tau}])^2}{\frac{1}{T} \sum_{t=1}^T (rx_{t,t+n}^{\tau} - \frac{1}{T} \sum_{\ell=1}^T rx_{t,t+n}^{\tau})^2}$, 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 two rows contain the R^2 from the regression of realized excess returns on the five yields and the five yields and five yield variances, respectively. Standard errors in parentheses are computed using the Hansen and Hodrick correction with number of lags equal to the number of overlapping months.

Regressing realized excess returns on model-implied expected excess in sample (1961-2014)

Maturity	Nonlinear model			$A_1(3)$ model			$A_0(3)$ model			5Y	5Y+5VAR			
	$\alpha \times 10^3$	β	R^2	FVE	$\alpha \times 10^3$	β	R^2	FVE	$\alpha \times 10^3$	β	R^2	FVE	R^2	R^2
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	-3.77 (3.92)	1.30 (0.45)	0.12	0.11	0.13	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	-7.35 (6.71)	1.35 (0.44)	0.14	0.12	0.14	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	-10.29 (8.71)	1.37 (0.41)	0.15	0.13	0.16	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	-12.61 (10.23)	1.37 (0.38)	0.17	0.15	0.17	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	-2.21 (7.99)	0.91 (0.55)	0.07	0.05	0.09	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	-5.19 (13.52)	0.95 (0.51)	0.09	0.07	0.10	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	-8.94 (17.40)	1.01 (0.47)	0.11	0.09	0.12	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	2.20 (11.54)	0.57 (0.60)	0.03	-0.02	0.07	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	0.43 (19.90)	0.68 (0.55)	0.05	0.01	0.08	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	1.82 (14.53)	0.59 (0.62)	0.04	-0.03	0.08	0.23

Table VI: *URP (Unspanned Risk Premia) regressions*. This table shows R^2 s (in percent) from regressions of excess returns on the five principal components (PCs) of yields. Panel A shows R^2 from regressions of one-year actual realized excess return on PCs of actual yields based on the sample 1961:07-2014:04. Panel B, C, and D show for each model in sample R^2 from regressions of model-implied one-year excess return on model-implied PCs of yields. Panel E, F, and G show for each model population R^2 s from regressions of realized one-year excess return on PCs of yields based on a simulated data sample of 1,000,000 months. The final column of Panels E-G shows the R^2 s when using the model-implied excess return instead of the model-implied PCs as independent variable.

Maturity	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1-PC_4	PC_1-PC_5	$E_t [rx_{t,t+1}^\tau]$
Panel A: R^2 in data (1961-2014)						
$\tau = 2$	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: In sample R^2 for nonlinear three-factor model						
$\tau = 2$	5.7	64.9	67.5	85.3	91.2	
$\tau = 3$	4.6	67.7	69.1	84.8	90.8	
$\tau = 4$	4.2	69.7	70.6	84.8	90.7	
$\tau = 5$	4.4	71.0	72.0	85.4	90.9	
Panel C: In sample R^2 for $A_1(3)$ model						
$\tau = 2$	10.8	99.8	100.0			
$\tau = 3$	10.5	99.7	100.0			
$\tau = 4$	10.2	99.6	100.0			
$\tau = 5$	9.9	99.5	100.0			
Panel D: In sample R^2 for $A_0(3)$ model						
$\tau = 2$	5.3	99.6	100.0			
$\tau = 3$	1.4	99.9	100.0			
$\tau = 4$	0.2	100.0	100.0			
$\tau = 5$	0.0	99.6	100.0			
Panel E: Population R^2 for nonlinear three-factor model						
$\tau = 2$	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 F: Population R^2 for $A_1(3)$ model						
$\tau = 2$	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
Panel G: Population R^2 for $A_0(3)$ model						
$\tau = 2$	0.4	9.5	9.6	9.6	9.6	9.6
$\tau = 3$	0.1	9.8	9.8	9.8	9.8	9.8
$\tau = 4$	0.0	10.6	10.6	10.6	10.6	10.6
$\tau = 5$	0.0	11.7	11.7	11.7	11.7	11.7

Table VII: *URP regression residuals and expected inflation.* We first run a regression of model-implied one year expected excess returns on the PCs of model-implied yields, $E_t [rx_{t,t+1}^{\tau}] = \alpha^{\tau,1:n} + \sum_{i=1}^n \beta^{\tau,1:n} PC_{i,t} + \varepsilon_t^{\tau,1:n}$, $n = 3, 4, 5$. Then we run a regression of the residual of this regression on expected inflation, measured by the cross-sectional average forecasts of the Michigan Survey of Consumers (π_t), $\varepsilon_t^{\tau,1:n} = \alpha^{\tau,n} + \beta^{\tau,n} \pi_t + \text{residual}$, $n = 3, 4, 5$. This table shows R^2 's (in percent), slope coefficient and the t-statistic from the second regression. Expected excess returns are measured as the expected one year bond 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 not available at monthly frequencies before 1978.

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 VIII: *Principal Component analysis of realized yield variances.* Principal components (PCs) are constructed from a panel of realized yield variances of constant-maturity zero-coupon bond yields with maturities ranging from one to five years. The contribution of the first PC, the first and second PC, and the first, second, and third PC to the total variation in the five realized yield variances are shown for the data, the nonlinear model, and the $A_1(3)$ model. Actual PC contributions are computed using monthly realized variance data (based on daily squared yield changes) from 1961:07 to 2014:04 obtained from Gurkaynak, Sack, and Wright (2007). Population PC contributions for the nonlinear and $A_1(3)$ model are computed using monthly realized variance data (based on daily squared yield changes) based on one simulated sample path of 1,000,000 months.

	PC_1	PC_1-PC_2	PC_1-PC_3
Data	0.9454	0.9922	0.9996
Nonlinear model	0.9750	0.9993	1.0000
$A_1(3)$ model	1.0000	1.0000	1.0000

Table IX: *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 in sample R^2 s for the nonlinear model from regressing model-implied instantaneous variance on the PCs of model-implied yields. Panel C shows in population 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) based on a sample of 1,000,000 simulated months. Panel D and E shows corresponding results for the $A_1(3)$ 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 residuals from the USV regressions in Panel A and B.

Maturity	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1-PC_4	PC_1-PC_5
Panel A: R^2 in the data (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: In sample R^2 for nonlinear three-factor model					
$\tau = 1$	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: Population R^2 for nonlinear three-factor model					
$\tau = 1$	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: In sample R^2 for $A_1(3)$ model					
$\tau = 1, \dots, 5$	21.5	22.3	100.0	100.0	100.0
Panel E: Population R^2 for $A_1(3)$ model					
$\tau = 1, \dots, 5$	0.0	0.0	45.8	45.8	45.8
Panel F: In sample PC analysis of USV regression residuals					
Data	91.8	98.7	99.9	100.0	100.0
Nonlinear model	97.9	99.9	100.0	100.0	100.0

Table X: *Principal Component analysis of yields*. Principal components (PCs) are constructed from a panel of constant-maturity zero-coupon bond yields with maturities ranging from one to five years. The contribution of the first PC, the first and second PC, and the first, second, and third PC to the total variation in the five bond yields are shown in Panel A. In Panel B yields for each bond are then regressed on the first three principal components and a constant (omitted). Actual PC contributions, slope coefficients, and R^2 s are computed using monthly data of one through five-year zero coupon bond yields from 1961:07 to 2014:04 obtained from Gurkaynak, Sack, and Wright (2007). For all three models population PC contributions, population slope coefficients, and population R^2 s are based on one simulated sample path of 1,000,000 months.

Panel A: Principal components of yields

	PC_1	PC_1-PC_2	PC_1-PC_3
Data	99.1909	99.9779	99.9996
Nonlinear model	99.6866	99.9977	100.0000
$A_1(3)$ model	99.9738	100.0000	100.0000
$A_0(3)$ model	99.3788	99.9819	100.0000

Panel B: Linearity in the cross-section of yields

Maturity	PC_1	PC_2	PC_3	R^2
Data (1961-2014)				
$\tau = 1$	0.47	-0.72	0.48	1.00
$\tau = 2$	0.46	-0.22	-0.52	1.00
$\tau = 3$	0.45	0.12	-0.46	1.00
$\tau = 4$	0.43	0.36	-0.02	1.00
$\tau = 5$	0.42	0.54	0.54	1.00
Nonlinear three-factor model in population				
$\tau = 1$	0.45	-0.67	0.52	1.00
$\tau = 2$	0.45	-0.29	-0.37	1.00
$\tau = 3$	0.45	0.04	-0.52	1.00
$\tau = 4$	0.45	0.33	-0.17	1.00
$\tau = 5$	0.44	0.59	0.54	1.00
$A_1(3)$ model in population				
$\tau = 1$	0.51	-0.66	0.53	1.00
$\tau = 2$	0.48	-0.21	-0.58	1.00
$\tau = 3$	0.44	0.14	-0.41	1.00
$\tau = 4$	0.41	0.39	0.04	1.00
$\tau = 5$	0.38	0.58	0.46	1.00
$A_0(3)$ model in population				
$\tau = 1$	0.47	-0.72	0.47	1.00
$\tau = 2$	0.46	-0.21	-0.52	1.00
$\tau = 3$	0.45	0.13	-0.46	1.00
$\tau = 4$	0.43	0.36	-0.01	1.00
$\tau = 5$	0.42	0.54	0.54	1.00

Table XI: *Parameter estimates of the one-factor nonlinear model.* This table contains parameter estimates and asymptotic standard errors (in parenthesis) for the nonlinear one factor model. The parameter estimates are based on yield and realized variance data for the sample period 1961:07-2014:04. The bond maturities range from 1-5 years and the data are obtained from Gurkaynak, Sack, and Wright (2007). The unscented Kalman filter is used to estimate the nonlinear model.

κ	ρ_0	ρ_X	λ_0	λ_X	γ	β	σ_y	σ_{rv}
0.04027 (0.03695)	0.03061 (0.06164)	0.01093 (0.0001309)	-0.6473 (0.5656)	0.05966 (0.03703)	0.01456 (0.03452)	-0.4206 (0.003205)	0.003122 (0.0004019)	0.0001671 (1.161e-05)

Table XII: *The cross-section of yields in one-factor models.* Principal components (PCs) are constructed from a panel of constant-maturity zero-coupon bond yields with maturities ranging from one to five years. The contribution of the first PC, the first and second PC, and the first, second, and third PC to the total variation in the five bond yields are shown in Panel A. In Panel B yields for each bond are then regressed on the first three principal components and a constant (omitted). Actual PC contributions, slope coefficients, and R^2 s are computed using monthly data of one through five-year zero coupon bond yields from 1961:07 to 2014:04. For the nonlinear one factor model and for the one factor Gaussian model population PC contributions, population slope coefficients, and population R^2 s are based on one simulated sample path of 1,000,000 months.

Panel A: Principal components of yields

	PC_1	PC_1-PC_2	PC_1-PC_3
Data	99.1909	99.9779	99.9996
Nonlinear model	99.9740	99.9999	100.0000
$A_0(1)$ model	100.0000	100.0000	100.0000

Panel B: Linearity in the cross-section of yields

Maturity	PC_1	PC_2	PC_3	R^2
Data (1961-2014)				
$\tau = 1$	0.47	-0.72	0.48	1.00
$\tau = 2$	0.46	-0.22	-0.52	1.00
$\tau = 3$	0.45	0.12	-0.46	1.00
$\tau = 4$	0.43	0.36	-0.02	1.00
$\tau = 5$	0.42	0.54	0.54	1.00
Nonlinear one factor Model in population				
$\tau = 1$	0.49	-0.65	0.50	1.00
$\tau = 2$	0.46	-0.26	-0.37	1.00
$\tau = 3$	0.44	0.08	-0.53	1.00
$\tau = 4$	0.43	0.37	-0.16	1.00
$\tau = 5$	0.41	0.60	0.56	1.00
Gaussian one factor model in population				
$\tau = 1$	0.47			1.00
$\tau = 2$	0.46			1.00
$\tau = 3$	0.45			1.00
$\tau = 4$	0.43			1.00
$\tau = 5$	0.42			1.00

Table XIII: *URP and USV regressions in the one-factor nonlinear model.* Panel A shows in sample R^2 s from regressions of model-implied one-year excess returns on the Principal Components (PCs) of model-implied yields. Panel B shows in sample R^2 s from regressing model-implied instantaneous variance on the PCs of model-implied yields. Model-implied PCs are constructed from a panel of constant-maturity zero-coupon bond yields with maturities ranging from one to five years. The in sample results are based on the sample period 1961:07-2014:04.

Maturity	PC_1	PC_1-PC_2	PC_1-PC_3	PC_1-PC_4	PC_1-PC_5
Panel A: In sample R^2 for nonlinear 1-factor model					
$\tau = 2$	13.0	69.3	98.1	98.4	100.0
$\tau = 3$	14.4	73.2	98.1	98.4	100.0
$\tau = 4$	16.7	76.7	98.1	98.4	100.0
$\tau = 5$	19.9	79.9	98.1	98.4	100.0
Panel B: In sample R^2 for nonlinear one factor model					
$\tau = 1$	52.0	64.8	96.1	96.8	99.9
$\tau = 2$	55.1	68.2	96.6	97.2	99.9
$\tau = 3$	57.9	71.3	97.0	97.5	100.0
$\tau = 4$	60.4	73.9	97.4	97.8	100.0
$\tau = 5$	62.7	76.2	97.7	98.0	100.0

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

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

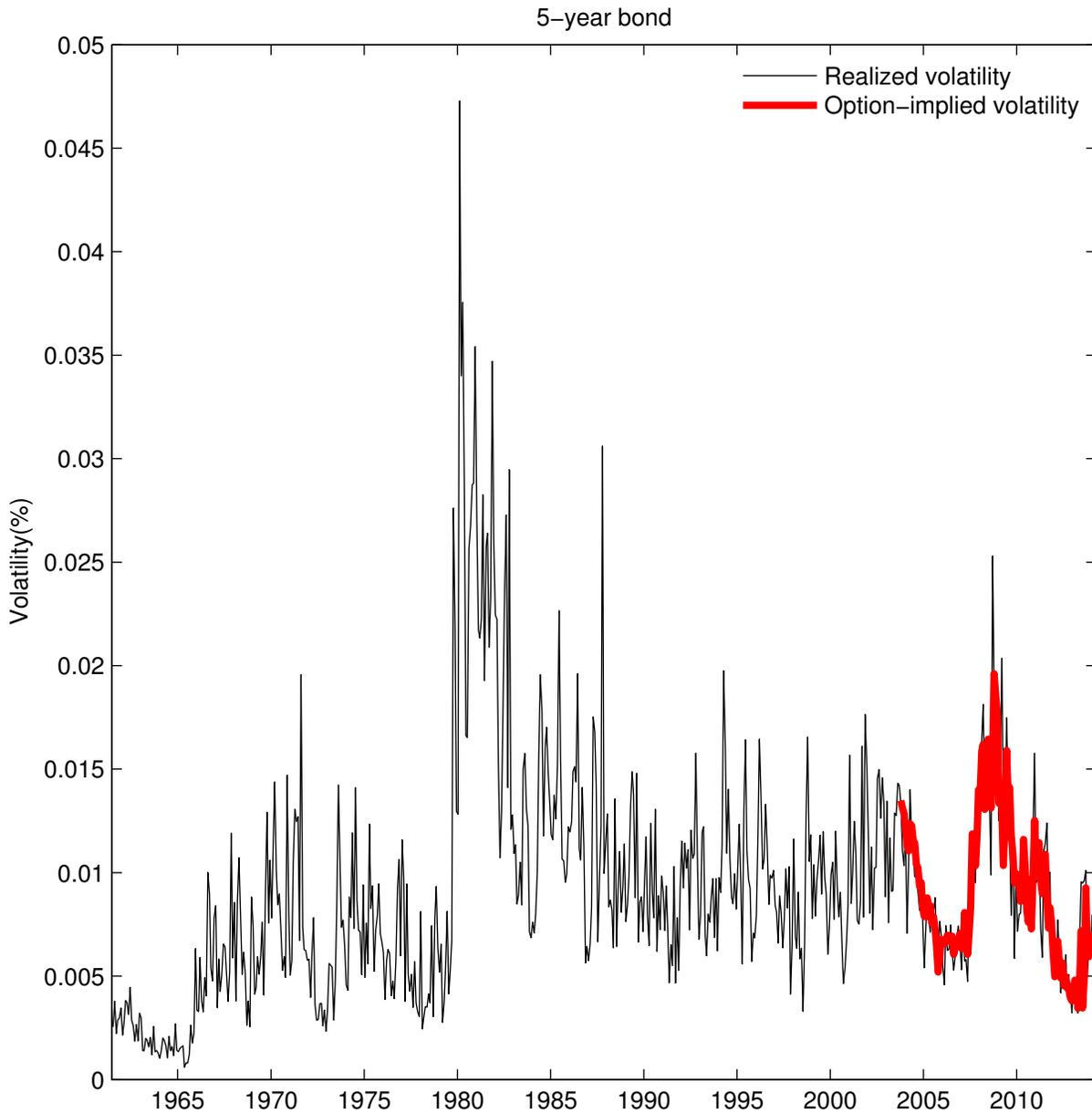


Figure 1: *Realized and option-implied yield volatility.* We use monthly estimates of realized yield variance based on daily squared yield changes. This graph shows that option-implied volatility tracks the realized volatility closely over the last 10 years (the correlation is 87%). Option-implied volatility 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.

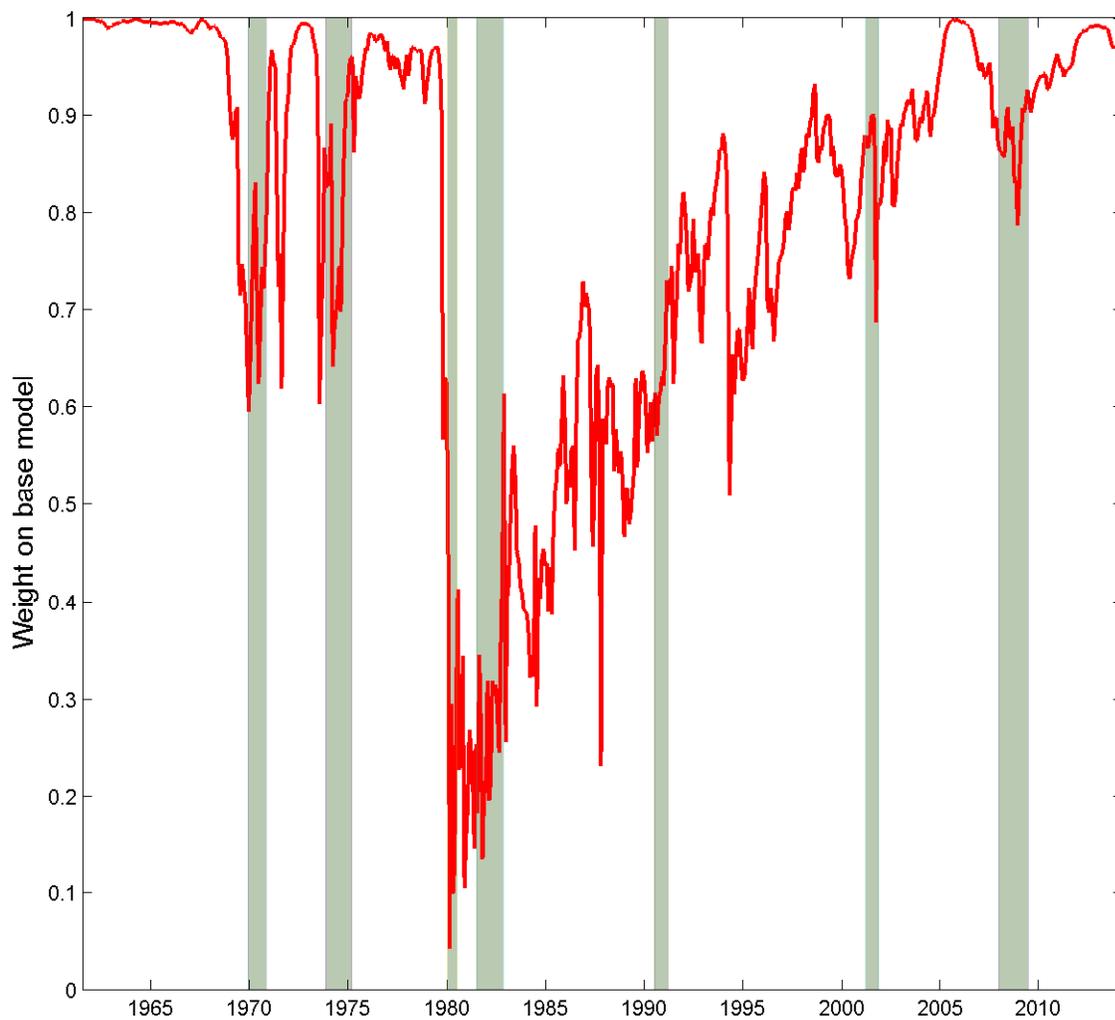


Figure 2: *Stochastic weight on Gaussian base model.* The bond price in the nonlinear model is $P(t, T) = s(t)P_0(t, T) + (1 - s(t))P_1(t, T)$ where $P_0(t, T)$ and $P_1(t, T)$ are bond prices that belong to the class of essentially affine Gaussian term structure models and $s(t)$ is a stochastic weight between 0 and 1. This figure shows the stochastic weight and the shaded areas show NBER recessions.

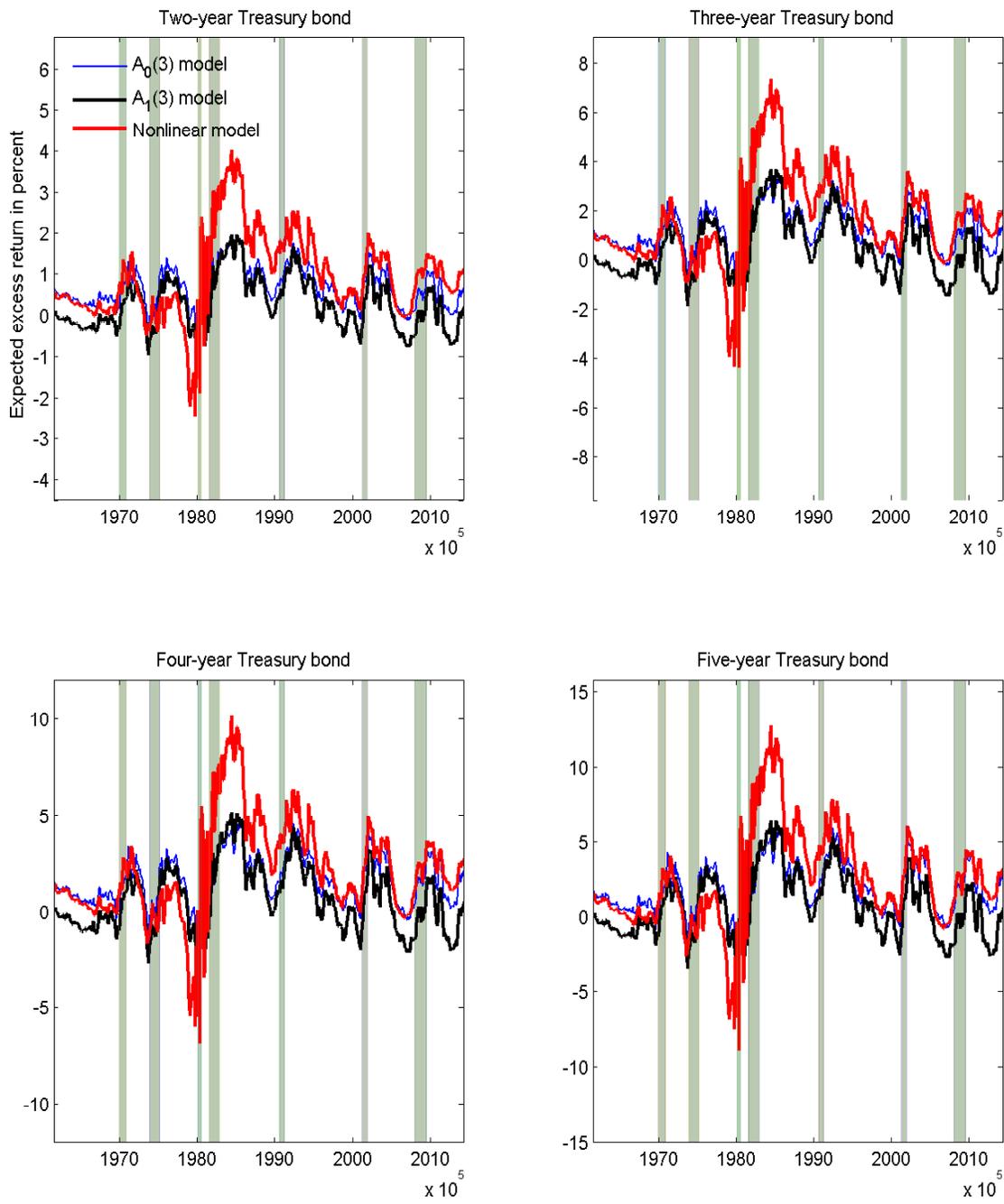


Figure 3: *Expected excess returns.* The graphs show the expected one year log excess returns of zero-coupon Treasury bonds with maturities of 2, 3, 4, and 5 years. The blue, black, and red lines show expected excess returns in the three-factor $A_0(3)$, $A_1(3)$, and nonlinear model, respectively. The shaded areas show NBER recessions.

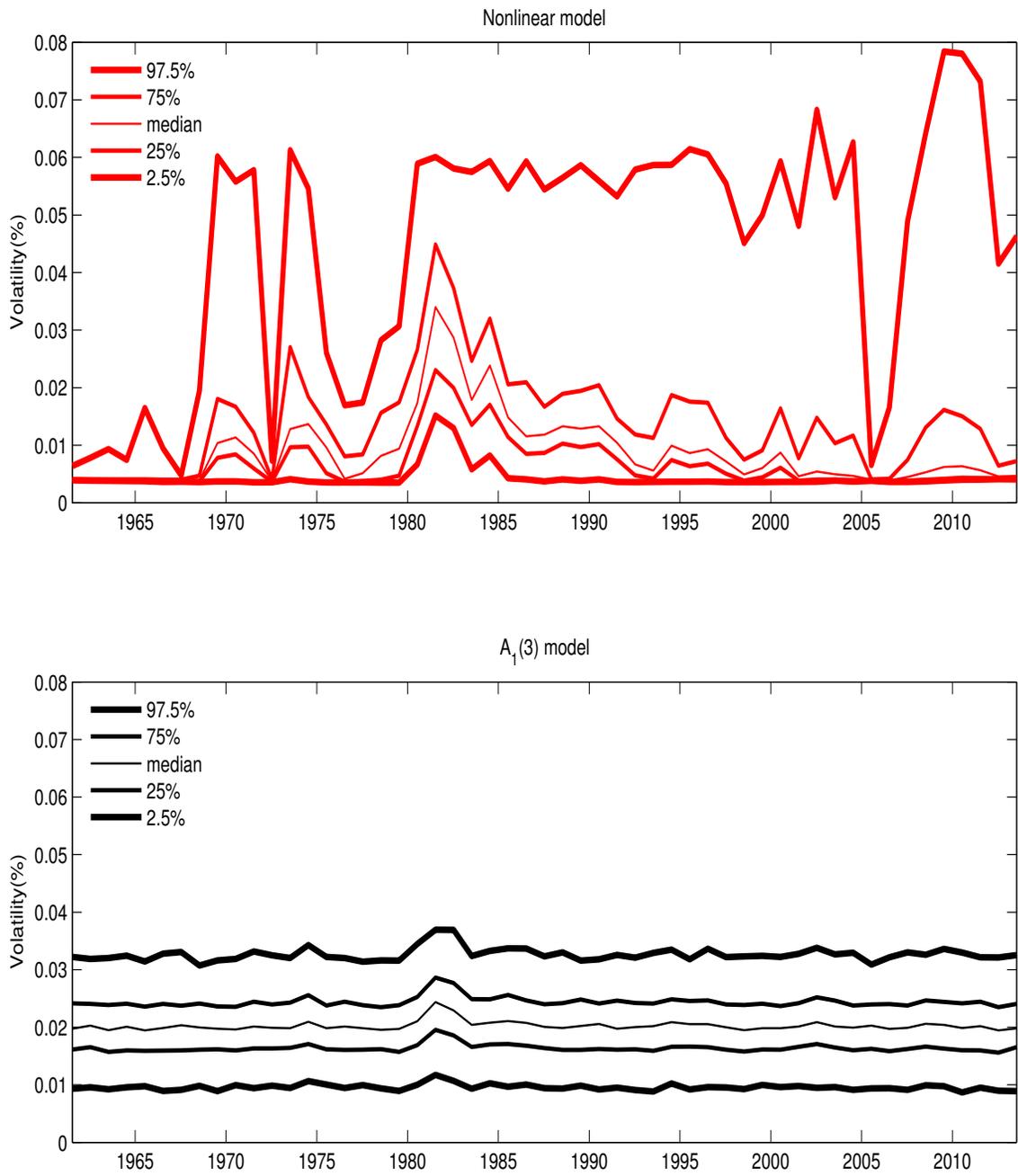


Figure 4: *Distribution of one-year ahead yield volatility.* The graphs show quantiles in the one-year ahead distribution of instantaneous volatility for the bond with a maturity of three years. The top graph shows the distribution in the three-factor nonlinear model, while the bottom graph shows the 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.

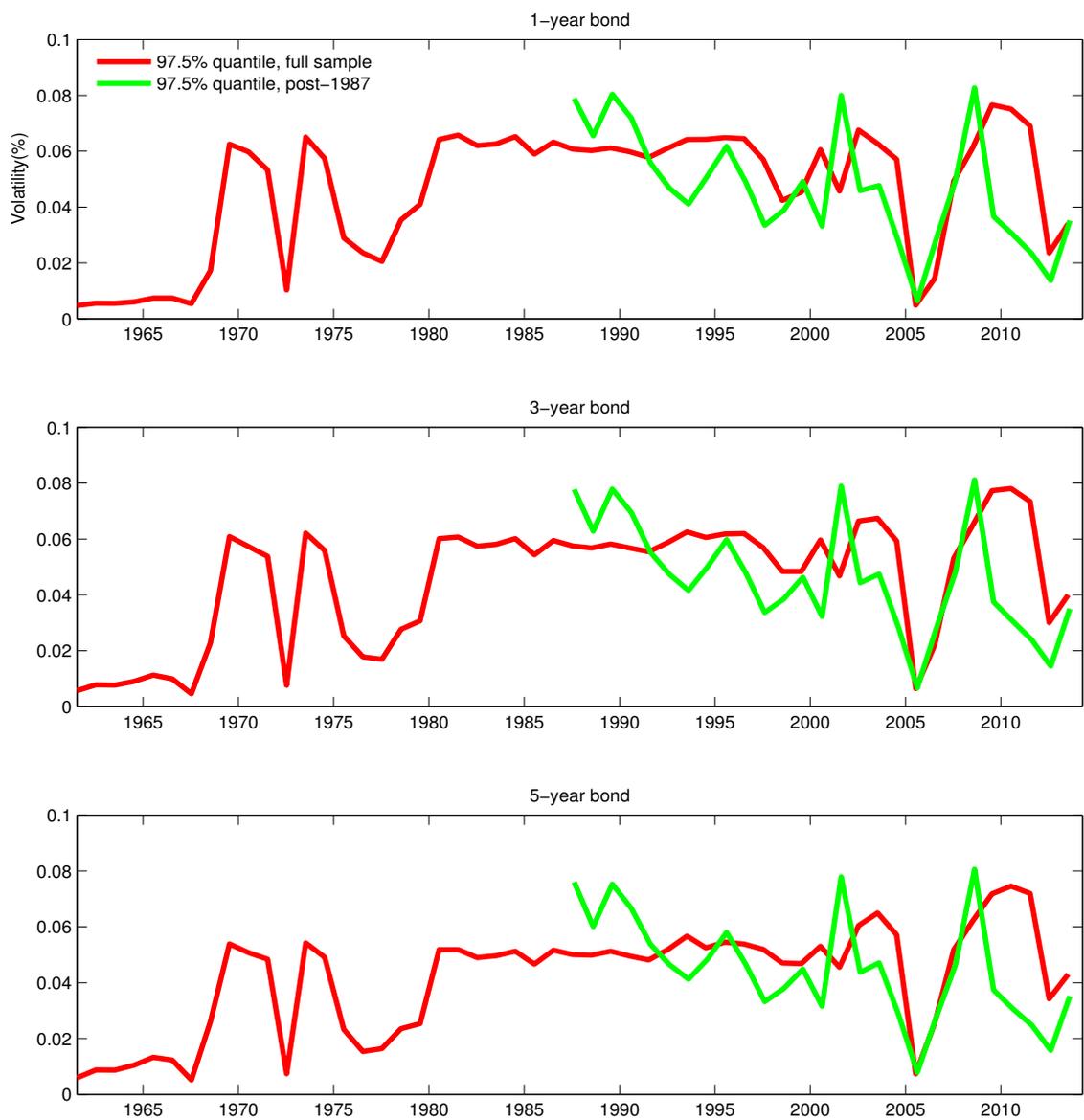


Figure 5: *Distribution of one-year ahead yield volatility for nonlinear model estimated using 1961-2014 and estimated using 1987-2014.* The graphs show the 97.5% quantiles in the one-year ahead distribution of instantaneous volatility. The red line shows the 97.5% quantiles in the three-factor nonlinear 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.

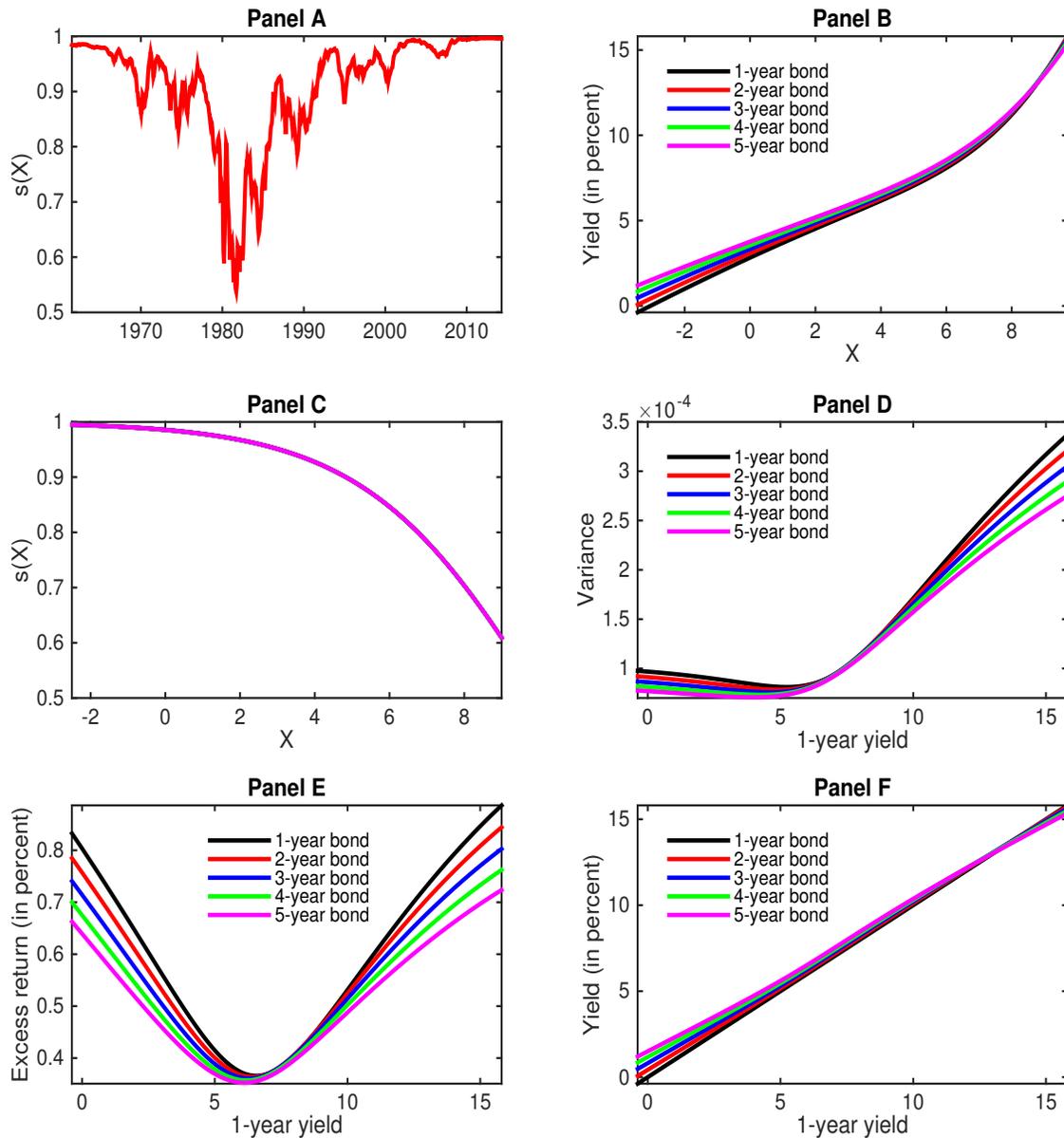


Figure 6: *Stochastic weight, yields, volatilities, and excess returns in a one-factor nonlinear model.* Panel A shows the estimated stochastic weight on the Gaussian base model for the sample period 1961-2014 and Panel C shows it as function of the factor X . Panel B and F show yields as function of the factor X and the one year yield, respectively. Panel D and E show yield variance and expected excess returns as a function of the one-year yield. The parameters for the one-factor nonlinear model are estimated using yields and realized yield variance of zero-coupon Treasury bonds with maturities ranging from one to five years. The range of X on the x-axis equals the range of X in the sample period 1961-2014.