RICHARD STARTZ
KWOK PING TSANG

An Unobserved Components Model of the Yield Curve

We develop an unobserved component model in which the short-term interest rate is composed of a stochastic trend and a stationary cycle. Using the Nelson–Siegel model of the yield curve as inspiration, we estimate an extremely parsimonious state-space model of interest rates across time and maturity. The time-series process suggests a specific functional form for the yield curve. We use the Kalman filter to estimate the time-series process jointly with observed yield curves, greatly improving empirical identification. Our stochastic process generates a three-factor model for the term structure. At the estimated parameters, trend and slope factors matter while the third factor is empirically unimportant. Our baseline model fits the yield curve well. Model generated estimates of uncertainty are positively correlated with estimated term premia. An extension of the model with regime switching identifies a high-variance regime and a low-variance regime, where the high-variance regime occurs rarely after the mid-1980s. The term premium is higher, and more so for yields of short maturities, in the high-variance regime than in the low-variance regime. The estimation results support our model as a simple and yet reliable framework for modeling the term structure.

JEL codes: C22, E32, E43
Keywords: term structure of interest rates, Nelson–Siegel yield curve, trend-cycle decomposition.

Some two decades ago, Charles Nelson and Andrew Siegel (Nelson and Siegel 1987) proposed a model of the term structure of interest in which yields across a cross-section of maturities are modeled as the integral from time 0 to maturity of the solution to a second-order differential equation. This Nelson–Siegel model has found wide practical application for modeling the yield curve at a point

Financial support from the Castor Professorship at the University of Washington is gratefully acknowledged. The authors thank Shelly Lundberg, Charles R. Nelson, two anonymous referees, and the editor, Pok-sang Lam, for helpful comments, and Robert Bliss for making his interest rate data available.

Richard Startz *is at the Department of Economics, University of Washington* (*E-mail:* startz@u.washington.edu). Kwok Ping Tsang *is at the Department of Economics, University of Washington* (*E-mail:* byront@vt.edu).

Received May 16, 2007; and accepted in revised form April 26, 2010.

Journal of Money, Credit and Banking, Vol. 42, No. 8 (December 2010)
© 2010 The Ohio State University

in time. Recent work by Diebold and Li (2006) shows that the Nelson–Siegel curve can be interpreted as maturity-varying factor loadings on three factors: level, slope, and curvature. This has opened the way for studying the time-series behavior of these factors as well as their relation to macroeconomic driving variables, which adds more economic content to the Nelson–Siegel curve.

Our approach might be summarized succinctly as we take Nelson–Siegel seriously.

Nelson–Siegel posit that the short rate follows a second-order differential equation. The solution to that equation gives future short rates, and integrating the solution gives current long rates under a no-premium expectations hypothesis. Because the functional form for the yield curve is flexible (while using very few parameters), it can fit yields across maturities quite well. However, nothing in these cross-section fits ties them to the original short-rate, time-series process. Here, in contrast, we create a tight empirical link between the short-rate process and yields across maturities, and in this way rejoin the underlying time-series and expectations hypothesis theory and the empirical yield curve.[1]

We start with a time-series model for the short rate, where the short rate is the sum of two unobserved components: a stochastic trend with unit root and a stationary cycle. This model has a univariate ARIMA representation that is a close stochastic analog to the Nelson–Siegel deterministic, second-order differential equation. We then integrate the future short-rate forecasts. Here, we make one significant departure from Nelson–Siegel. The Nelson–Siegel solution is based on a strong form (zero term premia) expectations hypothesis model of the term structure. Zero term premia are empirically untenable. Nonetheless, Nelson–Siegel cross-section fits are invariably made to observed yields, rather than observed yields adjusted for term premia. Our empirical model allows for nonzero premia, which is important in separating out expectations of future rates from the premium. In fact, we know that the weak form (nonzero, but time-invariant premia) expectations hypothesis is also violated in the data (Startz 1982, Campbell and Shiller 1991, Thornton 2006). In the final part of the paper, we introduce a Markov-switching model of a time-varying premium. While our model provides some suggestive evidence about the relation between uncertainty and a time-varying premium, the main contribution lies in the modeling of the expectations portion of long rates.

We combine the time-series model for the short rate with the equations giving the cross-section of yields at different maturities in a state-space model and estimate the underlying process. Because modeling the cross-sections adds very few parameters to the time-series model, a great deal of data are available to identify the parameters. This framework provides a good description of the dynamics of the short rate and also a satisfactory cross-sectional fit. While the resulting estimates allow for a level/slope/curvature factor interpretation, the decomposition into trend and cycle may be useful in relating interest rates to other trend/cycle macrodecompositions. We compare our estimates of trend and cycle to appropriate level and slope estimates

1. For a theoretically consistent approach tying together time series and cross-section in affine models, see De Jong (2000) and Christenson, Diebold, and Rudebusch (2009).

from unconstrained cross-section fits and find a close correspondence. Thus, once the time-constant parameters are estimated, our model can be used to draw current or future yield curves based only on knowledge of the short rate.

We also find an explanation for the fact that the "curvature factor" has been notoriously difficult to identify. In our three-factor model, the factors are trend, cycle, and lagged cycle. The third factor is identified only if the lagged cycle factor matters and the lagged cycle matters only if the cycle follows an AR(2). The time series on observed yield data is not powerful enough to clearly identify an AR(2) component. In contrast, our combined time-series/cross-section approach does give good identification.

Since our model takes uncertainty seriously and allows for term premia, we can offer some insight on the relation between risk and premia. While we make no attempt to build an optimizing model relating risk and return, the empirical relation we find between estimated premia and modeled uncertainty is interesting, as uncertainty is highly correlated with the level of the premium. We then take this further by allowing for Markov switching in variances, while retaining the linear unobserved component model for levels. We find evidence that there is Markov switching and that accounting for it improves the performance of the model, in particular Markov switching improves identification of the third factor in the Nelson–Siegel curve. Further, we find higher premia in higher variance states. Our estimates of the Markov states are highly significant in a Campbell–Shiller regression but do not eliminate the Campbell–Shiller paradox.

1. A SERIOUS, STOCHASTIC NELSON–SIEGEL MODEL

1.1 The Original Nelson–Siegel Model

In the original Nelson and Siegel (1987) article, the process for the short rate r (with maturity of one period) is assumed to be a nonhomogeneous second-order differential equation,

$$a\ddot{r} + b\dot{r} + cr = d. \tag{1}$$

Following the development in Nelson and Siegel (1987) and Diebold and Li (2006), and assuming that the roots of equation (1) are real and equal, the forward rate path, $r^f(m)$, as a function of maturity m, is

$$r^f(m) = \frac{d}{c} + \left(e - \frac{d}{c}\right)\exp\left(-m\frac{b}{2a}\right) + \left(\frac{2af}{b} + e - \frac{d}{c}\right)\frac{bm}{2a}\exp\left(-m\frac{b}{2a}\right), \tag{2}$$

where initial conditions are $r(0) = e$ and $\dot{r}(0) = f$.

Integrating the forward curve from 0 to m and dividing by m, we obtain the Nelson–Siegel yield curve $r^{(m)}$.

$$r^{(m)} = L + S\left(\frac{1-\exp(-m/\kappa)}{m/\kappa}\right) + C\left(\frac{1-\exp(-m/\kappa)}{m/\kappa} - \exp(m/\kappa)\right) \quad (3)$$

$$L = \frac{d}{c},\ S = e - \frac{d}{c} \text{ and } C = 2\left(f + e\sqrt{\frac{c}{a}} - \frac{d}{\sqrt{ac}}\right)\sqrt{\frac{a}{c}} - e + \frac{d}{c}. \quad (4)$$

The three coefficients in equation (4) are usually interpreted as the level, slope, and curvature factors.

In practice, equation (3) is usually estimated by period-by-period ordinary least squares (OLS), fixing the value of $1/\kappa$ at 0.0609. Any restrictions that might follow from equation (4) are ignored. The period t regression gives estimates of L_t, S_t, and C_t without imposing any time-series restriction, so in total $T \times 3$ structural parameters are estimated. These estimates are then collected as time series that summarize the dynamics of the term structure over time. The flexible functional form of equation (3) allows excellent fit of the term structure in cross sections—the R^2 is usually over 0.99. Recent cross-section estimates of Nelson–Siegel curves are given in Gürkaynak, Sack, and Wright (2006). Cross-section fits of no-arbitrage models, as surveyed in Duffee (2002), also produce excellent fits.

1.2 A Stochastic Model

Introducing an explicit stochastic specification in the form of a discrete time, unobserved components model, we write the short rate r_t as,

$$r_t = \tau_t + c_t \quad (5)$$

$$\tau_t = \tau_{t-1} + u_t,\ u_t \sim N\left(0, \sigma_u^2\right) \quad (6)$$

$$\phi(L)c_t = v_t,\ v_t \sim N\left(0, \sigma_v^2\right),\ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 \quad (7)$$

$$\operatorname{cov}(u_t, v_{t+k}) = \sigma_{uv} \text{ for } k = 0,\ \sigma_{uv} = 0 \text{ otherwise.} \quad (8)$$

In words, the short rate r_t is decomposed into a stochastic trend (i.e., a random walk) τ_t and an AR(2) cycle c_t, and the shocks to these two unobserved components are contemporaneously correlated. Note that equations (5)–(8) imply an ARIMA(2,1,2) univariate representation for the short rate in which the AR terms are ϕ_1 and ϕ_2 and the moving average (MA) terms depend on both ϕ_1 and ϕ_2 and the covariance matrix of the shocks to trend and cycle. Forward rates equal the current trend plus the forward forecast of the cycle.

Under the weak-form expectation hypothesis with m-term premium $\omega^{(m)}$, we can write out the first few yields as

$$
\begin{aligned}
r_t &= r_t^{(1)} = \tau_t + c_t, \\
r_t^{(2)} &= \omega^{(2)} + \frac{1}{2}(r_t + E_t r_{t+1}) = \omega^{(2)} + \tau_t + \frac{1+\phi_1}{2} c_t + \frac{\phi_2}{2} c_{t-1}, \\
r_t^{(3)} &= \omega^{(3)} + \frac{1}{3}(r_t + E_t r_{t+1} + E_t r_{t+2}) = \omega^{(3)} + \tau_t + \frac{1+\phi_1+\phi_1^2+\phi_2}{3} c_t \\
&\qquad + \frac{\phi_2 + \phi_1\phi_2}{3} c_{t-1},
\end{aligned} \tag{9}
$$

and so on. Note in general we can write:

$$
r_t^{(m)} = \omega^{(m)} + \tau_t + f(\phi_1, \phi_2, m) c_t + g(\phi_1, \phi_2, m) c_{t-1}, \tag{10}
$$

where the functions $f(.)$ and $g(.)$ are the factor loadings. The factor loadings can be written as functions of the inverse roots of the AR polynomial (when the roots are real), as shown in equation (11). As a convenience, we provide the limiting case of equal roots in equation (12).

$$
\begin{aligned}
r_t^{(m)} &= \omega^{(m)} + \tau_t + \left(\frac{1}{\eta_1 - \eta_2}\right)\left(\eta_1\left(\frac{1-\eta_1^m}{m(1-\eta_1)}\right) - \eta_2\left(\frac{1-\eta_2^m}{m(1-\eta_2)}\right)\right) c_t \\
&\quad + \left(\frac{\eta_1\eta_2}{\eta_1 - \eta_2}\right)\left(\left(\frac{1-\eta_2^m}{m(1-\eta_2)} - \left(\frac{1-\eta_1^m}{m(1-\eta_1)}\right)\right)\right) c_{t-1}
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
r_t^{(m)} &= \omega^{(m)} + \tau_t + \frac{1}{m}\left[\frac{1-\eta^m}{1-\eta} + \frac{\eta - m\eta^m + (m-1)\eta^{m+1}}{(1-\eta)^2}\right] c_t \\
&\quad - \frac{\eta}{m}\left[\frac{\eta - m\eta^m + (m-1)\eta^{m+1}}{(1-\eta)^2}\right] c_{t-1}.
\end{aligned} \tag{12}
$$

Equation (10) offers a three-factor model. The first factor is the trend, which is the same as the level factor identified in the literature. The second factor is the cycle, which is the slope using the definition $r_t^{(\infty)} - r_t^{(1)} - \omega^{(\infty)}$. Note that unlike in the usual empirical implementation, we separate term premia out from the slope definition. The third factor is the lagged cycle. The lagged cycle does not correspond to a curvature factor.

We are going to argue that for recent American data, only level and slope factors are well identified. The form of the time-series process, together with equations (11) and (12), determines the factor loadings. If the short-rate process was a pure random walk, then only the level factor would exist and the yield curve (aside from premia) would be a horizontal line. Slightly more generally, suppose that the short-rate process consisted of a random walk plus white noise; in our formulation suppose

TABLE 1

ARIMA Representations of Unobserved Component Models of the 1-Month Rate

	Random walk	Trend plus white noise cycle	Trend plus AR(1) cycle	Trend plus AR(2) cycle
ϕ_1			−0.441	0.860
			(0.486)	(0.373)
ϕ_2				−0.217
				(0.314)
θ_1		0.072	0.515	−0.807
		(0.052)	(0.464)	(0.380)
θ_2				0.087
				(0.329)
Log-likelihood	−378.41	−377.58	−377.14	−374.63
R^2	0	0.004	0.007	0.020

that $\phi_1 = \phi_2 = 0$. By inspection of equation (12) with $\eta = 0$, we see that the factor loading on the lagged cycle equals 0 and the factor loading on the slope equals $1/m$. The yield curve would at all times be a hyperbola. Neither horizontal line nor hyperbola seems sufficiently flexible to describe historically observed yield curves.

If the cyclical component of the short rate follows an AR(1) then the lagged cycle drops out of equation (11) as one of the roots equals zero. We have a two-factor model with slope loading $\frac{1-\phi_1^m}{m(1-\phi_1)}$. This model permits quite flexible yield curves, but not ones with hump shapes. (Negative ϕ_1 permits an oscillating slope load, which is not likely to be of much relevance for observed yield curves). In our data, a hump of more than 5 basis points occurs less than 5% of the time. Finally, a short-rate process where $\phi_2 \neq 0$ allows for hump shapes as well.

Our empirical estimates combine time-series and cross-section information in a way that provides good identification of both ϕ_1 and ϕ_2, including estimates that ϕ_2 is small—so that if there is a third factor its loading has been quite small in recent U.S. history. In addition, our evidence suggests that the equal root model of equation (12), which is similar to the models commonly found in the literature, would be better eschewed as empirically one (inverse) root is large while the other is close to zero. As a benchmark, ARIMA representation time-series results for the short rate are given in Table 1 (asymptotic standard errors in parentheses).

As is evident from Table 1, the short rate looks a great deal like a random walk. While the t-statistic on ϕ_1 in the ARIMA(2,1,2) model is significant at the 5% level, the confidence interval tells us little other than that ϕ_1 is probably positive. The confidence interval for ϕ_2 includes pretty much all interesting values. The p-value for the likelihood-ratio test of ARIMA(2,1,2) versus a random walk is 0.11. In summary, the time-series evidence suggests some role for the slope factor, although one that is difficult to pin down, and tells us essentially nothing about loadings on a third factor.

We turn now to a state-space representation that incorporates both time series and cross-section information. Let the M-vector of interest rates at time t be $\mathbf{r}_t = [r_t^{(1)} r_t^{(3)}, \ldots, r_t^{(120)}]'$. We augment the time-series model for the short rate given in equations (5)–(8) with the yield curves in equation (13).

$$\mathbf{r}_t = \Omega + \mathbf{I}\tau_t + \mathbf{F}(\phi_1, \phi_2)c_t + \mathbf{G}(\phi_1, \phi_2)c_{t-1} + \boldsymbol{\varepsilon}_t,\ \ \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \Sigma), \tag{13}$$

where the vector Ω contains the constant term premia for yields with maturity of more than 1 month, and the vector $\boldsymbol{\varepsilon}_t$contains the errors for all yields except the 1-month yield (i.e., the 1-month yield is restricted as being estimated perfectly). We follow the universal convention in the term structure literature of treating these as measurement errors, although the reader skeptical of there being significant difficulty in measuring returns to U.S. Treasury securities might think $\boldsymbol{\varepsilon}$ more a measure of how well our model fits the data. We assume the covariance–variance matrix for the measurement errors, Σ, to be diagonal, except for the 1-month yield whose measurement error is always zero. The measurement errors $\boldsymbol{\varepsilon}_t$ are assumed to be uncorrelated with the two state variable shocks u_t and v_t. The loadings $f(.)$ and $g(.)$ are functions of the two AR(2) coefficients.

Our state space model imposes a stringent constraint on the structural parameters. There are $M \times T$ observations on yields. The usual repeated cross-section approach explains the observations with $T \times 3 + 1$ parameters, three factors each period and κ. Our model requires four ARMA parameters plus M term premia. (A second model, below, adds another $M + 2$ parameters). Note that the three-latent factor model of Diebold, Rudebusch, and Aruoba (2006) effectively uses only three means and 3^2 VAR(1) parameters.

2. EMPIRICAL ESTIMATES

2.1 Data

Our observations come from the same data set used in Diebold and Li[2] of monthly zero-coupon CRSP Treasury data from January 1970 to December 2000 for maturities 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. We use unsmoothed Fama–Bliss[3] yields, and bonds with option features and special liquidity problem are eliminated. Figure 1 is a picture of the short (1-month) rate. Table 2 provides the summary statistics. The average premium of the 10-year rate over the 1 month is 160 or 190 basis points, depending on whether one uses mean or median. Standard deviations of yields decline modestly with maturity.

According to our model, long-maturity yields load almost entirely on a random walk trend while short-maturity yields include a stationary component as well. Cochrane's (1988) variance ratio can tell us whether this is a reasonable characterization of the data. The variance ratio for horizon k, $R_k = V_k/V_1$, where $V_k = \mathrm{var}(r_{t+k}^{(m)} - r_t^{(m)})/k$ tells us the fraction of the monthly variation in the yield that is due to permanent shocks. As k increases, the ratio R_k should stay at one if the yield

2. The data set is an extract of data supplied by Robert Bliss (see Bliss 1997).

3. See Fama and Bliss (1987) and Bliss (1997) for a discussion of the method.

FIG. 1. One-Month Treasury Bill Yields.

TABLE 2

DESCRIPTIVE STATISTICS FOR THE YIELD DATA

Maturity	Mean	Median	Maximum	Minimum	SD
1	6.44	5.69	16.16	2.69	2.58
3	6.75	5.93	16.02	2.73	2.66
6	6.98	6.24	16.48	2.89	2.66
9	7.1	6.4	16.39	2.98	2.64
12	7.2	6.61	15.82	3.11	2.57
15	7.31	6.75	16.04	3.29	2.52
18	7.38	6.78	16.23	3.48	2.5
21	7.44	6.81	16.18	3.64	2.49
24	7.46	6.81	15.65	3.78	2.44
30	7.55	6.93	15.4	4.04	2.36
36	7.63	7.06	15.77	4.2	2.34
48	7.77	7.22	15.82	4.31	2.28
60	7.84	7.37	15.01	4.35	2.25
72	7.96	7.42	14.98	4.38	2.22
84	7.99	7.45	14.98	4.35	2.18
96	8.05	7.51	14.94	4.43	2.17
108	8.08	7.54	15.02	4.43	2.18
120	8.05	7.59	14.93	4.44	2.14

follows a pure random walk, and the ratio R_k should decline toward zero if the yield r_t is trend stationary. In Table 3, we see the variance ratio R_k decreases with horizon for shorter yields but that for the longer yields stays around one. This is consistent with our model.

TABLE 3

COCHRANE VARIANCE RATIO FOR SIX YIELDS

Horizon in months	Maturity					
	1-month	3-month	12-month	24-month	48-month	120-month
2	1.058	1.120	1.148	1.173	1.079	1.100
3	1.028	1.110	1.129	1.161	1.045	1.110
4	0.981	1.082	1.076	1.105	1.015	1.093
5	0.929	1.037	0.999	1.040	0.966	1.087
6	0.890	1.009	0.971	1.009	0.952	1.092
7	0.833	0.951	0.933	0.982	0.949	1.104
8	0.768	0.877	0.864	0.929	0.918	1.104
9	0.751	0.848	0.830	0.902	0.906	1.100
10	0.746	0.846	0.811	0.887	0.900	1.104
11	0.752	0.863	0.803	0.875	0.897	1.123
12	0.768	0.888	0.821	0.893	0.918	1.150
24	0.753	0.915	0.822	0.853	0.856	1.007
28	0.712	0.864	0.786	0.807	0.816	0.948
120	0.359	0.453	0.486	0.601	0.714	0.965

2.2 Estimation of the Baseline Model

We estimate the model equations (5)–(8) and equation (13) by Gibbs sampling organized in two steps. Given the parameters $\{\Sigma, \Omega, \sigma_u^2, \sigma_v^2, \sigma_{uv}, \phi_1, \phi_2\}$ and the data, we run a Kalman filter to generate cycle and trend. Given trend and cycle, we use Gibbs sampling to draw new values of the parameters. Details are in an unpublished appendix available upon request.

For the sampler, we choose diffuse (improper) priors for all the variance elements and normal, but effectively diffuse, priors for the premia with means set at the sample average difference $r^{(m)} - r^{(1)}$ and standard deviations set at 10. Because in unobserved components (UC) models there is always an identification issue in ensuring that the estimated cycle is stationary, we use persistent but stationary priors for the AR coefficients, $N(\left[\begin{smallmatrix} 1.1 \\ -.2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right])$. Because draws of the AR coefficients with near-unit roots tend either be followed by nonstationary draws or long sequences with near-unit roots, we use rejection sampling and discard draws where $\tilde{\phi}_1 + \tilde{\phi}_2 \geq 0.95$. We run the Gibbs sampler 10,000 times, discarding the first 2,000 draws.

Posterior means, standard deviations, and 95% confidence bands appear in Tables 4–6. The cycle variance is estimated to be something over twice the size of the trend variance, and the two shocks are mildly negatively correlated with an estimated correlation coefficient of –0.31. The mean estimate of ϕ_1 happens to be the same as the time-series-only estimate, 0.86, but it is now precisely identified with 95% band (0.768, 0.952) as compared (0.13, 1.59). Where ϕ_2 was essentially unidentified in the time-series estimate, the Gibbs estimate is probably positive but certainly small with 95% band (-0.050, 0.134). Estimates of term premia rise with maturity, are essentially identical—within 3 basis points—of mean differences in yields, and are tightly estimated with standard deviations between 2 and 8 basis points.

TABLE 4

GIBBS SAMPLING RESULTS: COVARIANCE MATRIX FOR THE STATE VARIABLES, THE AR(2) PARAMETERS, AND THE TRANSITIONAL PROBABILITIES

	Posterior for baseline model				Posterior for model with switching			
	Mean	SD	95% bands		Mean	SD	95% bands	
Trend variance	0.196	0.019	0.168	0.232				
Cycle variance	0.466	0.061	0.384	0.584				
Covariance	−0.104	0.037	−0.175	−0.055				
Low trend variance					0.156	0.019	0.127	0.191
Low cycle variance					0.240	0.033	0.191	0.299
Low covariance					−0.098	0.024	−0.140	−0.061
High trend variance					0.281	0.068	0.186	0.403
High cycle variance					1.034	0.253	0.679	1.499
High covariance					−0.198	0.122	−0.420	−0.023
phi 1	0.861	0.054	0.771	0.950	0.823	0.039	0.753	0.879
phi 2	0.038	0.054	−0.052	0.129	0.109	0.035	0.062	0.177
q (low var prob)					0.869	0.023	0.829	0.905
p (high var prob)					0.796	0.033	0.740	0.848

TABLE 5

GIBBS SAMPLING RESULTS: VARIANCE OF THE MEASUREMENT ERRORS

Maturity	Posterior for baseline model					Posterior for model with switching				
	Mean	SD	95% bands		R^2	Mean	SD	95% bands		R^2
3	0.076	0.007	0.065	0.089	0.989	0.053	0.004	0.046	0.060	0.993
6	0.154	0.017	0.128	0.185	0.978	0.074	0.007	0.063	0.086	0.990
9	0.191	0.024	0.157	0.235	0.973	0.081	0.008	0.069	0.095	0.988
12	0.180	0.024	0.148	0.224	0.973	0.073	0.007	0.063	0.085	0.989
15	0.149	0.017	0.126	0.182	0.976	0.064	0.005	0.056	0.073	0.990
18	0.137	0.016	0.115	0.166	0.978	0.061	0.005	0.053	0.070	0.990
21	0.120	0.015	0.100	0.148	0.981	0.049	0.004	0.043	0.057	0.992
24	0.100	0.013	0.082	0.125	0.983	0.040	0.004	0.034	0.046	0.993
30	0.060	0.007	0.051	0.072	0.989	0.029	0.003	0.025	0.035	0.995
36	0.037	0.004	0.032	0.045	0.993	0.018	0.002	0.015	0.021	0.997
48	0.013	0.001	0.011	0.015	0.998	0.009	0.001	0.008	0.011	0.998
60	0.003	0.001	0.002	0.004	0.999	0.005	0.001	0.004	0.006	0.999
72	0.020	0.002	0.016	0.024	0.996	0.015	0.002	0.013	0.018	0.997
84	0.035	0.004	0.030	0.041	0.993	0.029	0.003	0.025	0.034	0.994
96	0.056	0.006	0.046	0.067	0.988	0.038	0.004	0.032	0.045	0.992
108	0.071	0.008	0.058	0.085	0.985	0.050	0.005	0.042	0.058	0.990
120	0.096	0.009	0.082	0.111	0.979	0.079	0.007	0.068	0.092	0.983

TABLE 6

Gibbs Sampling Results: Term Premia

	Posterior for baseline model				Posterior for model with switching							
					Low-variance regime				High-variance regime			
Maturity	Mean	SD	95% bands		Mean	SD	95% bands		Mean	SD	95% bands	
3	0.196	0.020	0.164	0.230	0.307	0.016	0.281	0.333	0.490	0.024	0.453	0.531
6	0.317	0.034	0.265	0.377	0.532	0.026	0.489	0.575	0.889	0.038	0.830	0.955
9	0.396	0.048	0.329	0.484	0.652	0.033	0.597	0.706	1.079	0.052	1.005	1.172
12	0.485	0.060	0.402	0.596	0.745	0.038	0.682	0.807	1.189	0.064	1.100	1.301
15	0.600	0.072	0.503	0.734	0.849	0.042	0.779	0.916	1.281	0.073	1.182	1.411
18	0.679	0.082	0.570	0.831	0.920	0.045	0.846	0.992	1.345	0.084	1.234	1.489
21	0.745	0.091	0.626	0.913	0.982	0.048	0.903	1.060	1.403	0.092	1.284	1.561
24	0.775	0.099	0.648	0.960	0.997	0.051	0.913	1.078	1.401	0.101	1.272	1.573
30	0.905	0.113	0.761	1.117	1.089	0.054	0.999	1.173	1.438	0.116	1.293	1.633
36	1.001	0.125	0.841	1.233	1.166	0.058	1.071	1.258	1.492	0.126	1.335	1.703
48	1.189	0.143	1.006	1.456	1.302	0.062	1.200	1.399	1.553	0.145	1.376	1.793
60	1.288	0.155	1.089	1.577	1.373	0.066	1.265	1.473	1.582	0.159	1.391	1.843
72	1.446	0.165	1.233	1.752	1.489	0.069	1.377	1.595	1.635	0.168	1.431	1.916
84	1.484	0.172	1.262	1.801	1.520	0.071	1.405	1.628	1.657	0.175	1.445	1.951
96	1.578	0.176	1.351	1.909	1.578	0.072	1.460	1.692	1.661	0.183	1.439	1.968
108	1.606	0.182	1.368	1.943	1.611	0.074	1.489	1.726	1.701	0.185	1.474	2.014
120	1.582	0.186	1.339	1.928	1.579	0.075	1.456	1.698	1.659	0.189	1.426	1.983

The model fits well, although as would be expected not so well as unconstrained cross-section fits. The variances of the "measurement" errors show that the model has some difficulty fitting yields of shorter maturities in particular. For example, the standard deviation of the measurement error for the 9-month yield, the largest reported, is 44 basis points, corresponding to an R^2 of only 0.973.

Figures 2a and 2b show mean estimates of trend and cycle from the Gibbs sampler together with trend and slope factors estimated from unconstrained cross-section regressions. The estimates are similar to one another except in the early 1990s, when we find lower trend and higher cycle than appears in cross-section estimates. To compare with cross-section estimates, we add in the estimated 120-month term premium.

Figure 3 shows slope (solid line) and lagged cycle (dotted line) loadings at mean Gibbs estimates. The slope loading is relatively large at least at maturities of a few years. The lagged cycle loading is small. Since we estimate the cycle variance to be considerably higher than the slope variance, the estimated slope loading is large enough that the slope factor plays an important role in setting the yield curve.

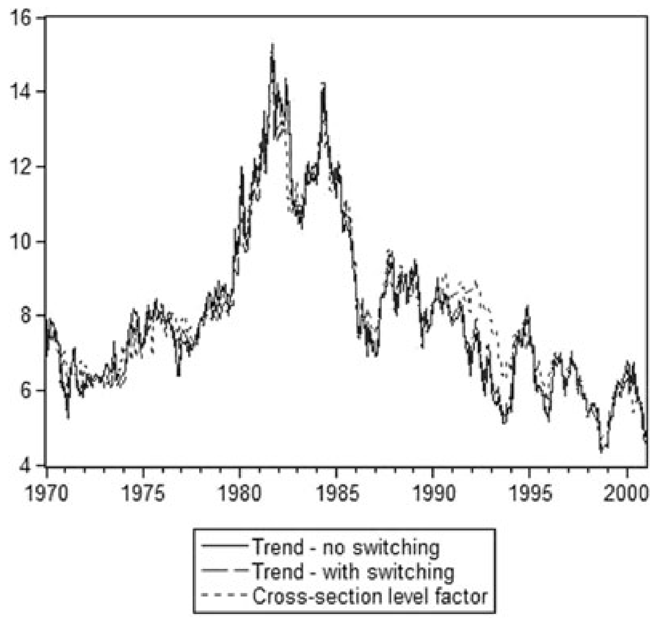
The relation between the premia and risk calculations from our model is intriguing. The surprise in the yield $r_t^{(k)}$ is

$$\begin{aligned} r_t^{(m)} - \mathrm{E}_{t-1}\big(r_t^{(m)}\big) &= \omega^{(m)} - \mathrm{E}_{t-1}\omega^{(m)} + \tau_t - \mathrm{E}_{t-1}\tau_t + f(m)(c_t - \mathrm{E}_{t-1}c_t) \\ &\quad + g(m)(c_{t-1} - \mathrm{E}_{t-1}c_{t-1}) \\ &= u_t + f(m)v_t. \end{aligned} \tag{14}$$

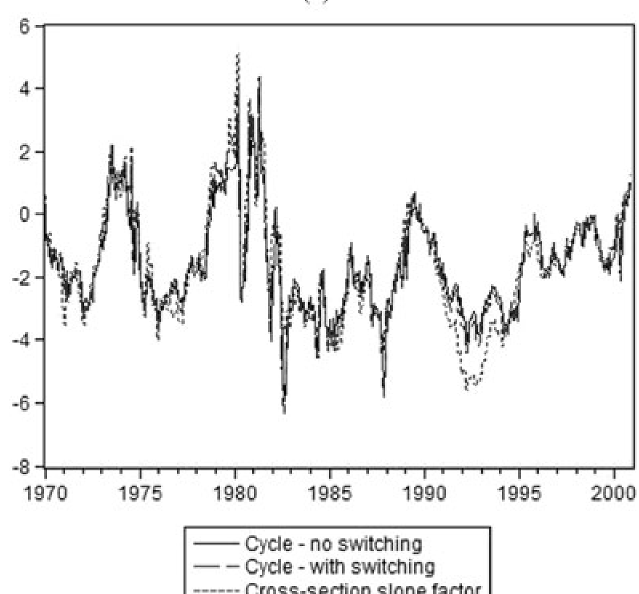
The conditional variance is a function of the model parameters, $\mathrm{var}(r^{(m)}) = \sigma_u^2 + f^2(m)\sigma_v^2 + 2f(m)\sigma_{uv}$. Since the price of a bond is $p^{(m)} = 1/(1 + r^{(m)})^m$, the usual Taylor series approximation gives $\mathrm{var}((p^{(m)} - p_{t-1}^m)/p_{t-1}^m) \approx m^2\mathrm{var}(r^{(m)})$. We compute this variance using the estimated parameters. Figure 4 shows the relation between estimated term premia and estimated uncertainty (for an $m-1$ period bond). Visually quite strong, the correlation across maturities between the premium and standard deviation is 0.94. Regressing premia on the standard deviation, despite the fact the relation is not quite linear, shows that a 100 basis point increase in standard deviation gives a 3 basis point increase in the premium.

2.3 Estimation of a Regime-Switching Model

Inspection of Figure 1 suggests that volatility of the short rate changes over time. Evidence in the last section suggests a link between uncertainty and premia, and the literature certainly suggests that premia are time varying. We build changing volatility into the model without interfering with the link between the time-series process and cross-sections in the model by allowing for regime switching in the variance of trend and cycle shocks and in the term premia, while holding the AR parameters in common. This means that factors shift between regimes but factor loadings do not. Letting $S = 1$ denote the high-variance regime, the model becomes



(a)



(b)

FIG. 2. (a) Estimated Trend versus Estimated Cross-Section Level Factors. (b) Estimated Cycle versus Estimated Cross-Section Slope Factors.

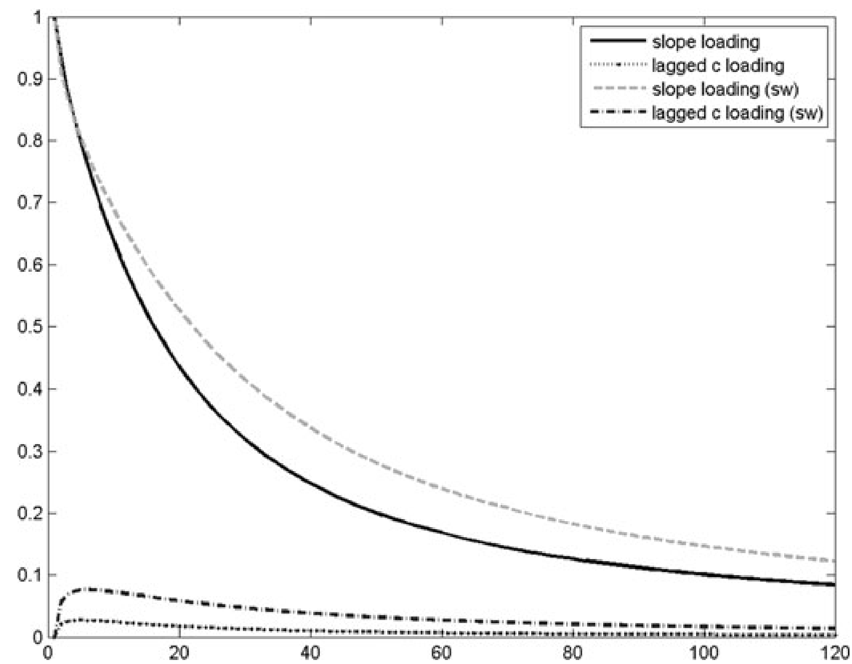


FIG. 3. Slope and Lagged Cycle Loadings.

$$\mathbf{r}_t = \omega_0(1 - S_t) + \omega_1 S_t + \mathbf{I}\tau_t + \mathbf{F}(\phi_1, \phi_2)c_t + \mathbf{G}(\phi_1, \phi_2)c_{t-1} + \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim N(0, \Sigma) \tag{15}$$

$$\tau_t = \tau_{t-1} + u_t,\ u_t \sim N\left(0, \sigma_{u0}^2(1 - S_t) + \sigma_{u1}^2 S_t\right) \tag{16}$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + v_t,\ v_t \sim N\left(0, \sigma_{v0}^2(1 - S_t) + \sigma_{v1}^2 S_t\right) \tag{17}$$

$$\text{cov}(u_t, v_t) = \sigma_{uv0}(1 - S_t) + \sigma_{uv1} S_t, \tag{18}$$

where S_t evolves as a two-state, first-order Markov-switching process with transitional probabilities

$$\Pr[S_t = 0 \mid S_{t-1} = 0] = q \text{ and } \Pr[S_t = 1 \mid S_{t-1} = 1] = p. \tag{19}$$

We use the same priors as above, with the addition of normal mean 0.9 and standard deviation 1.0 priors for p and q. Table 4 shows the posterior distributions. Both states are fairly persistent, with the expected duration of the low-variance state

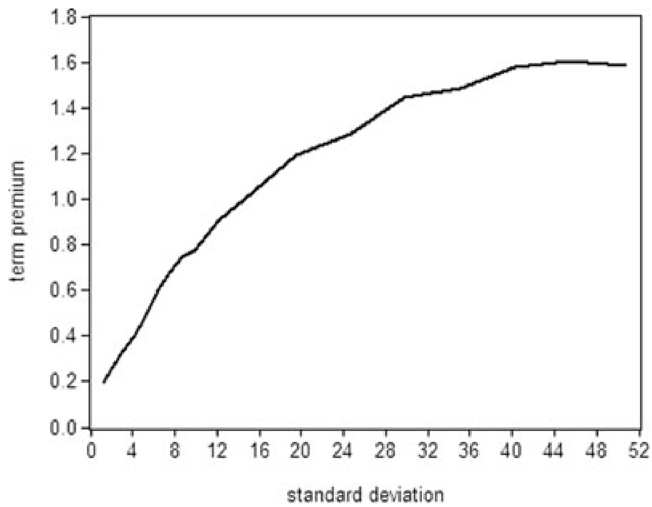


FIG. 4. Term Premia versus Conditional Standard Deviation of Prices.

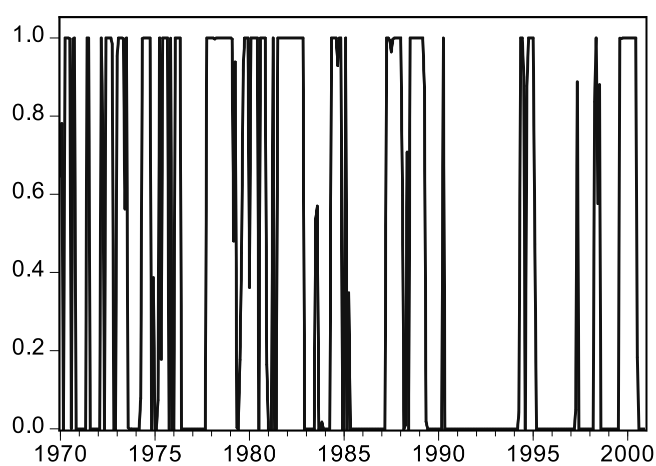


FIG. 5. Probability of High-Variance Regime.

(7.6 months) being greater than that of the high-variance state (4.6 months). Figure 4 plots the posterior probability of being in the high-variance regime. Before the 1980s, the yield curves switches frequently between the two regimes, and during the Volcker disinflation the yield curve is mainly in the high-variance state. From the mid-1980s onward the yield curve is mainly in the low-variance regime, with some short and infrequent spells of high variance.

In contrast to the earlier results, ϕ_2 and a third factor are now well identified. The loading on the third factor remains small. Both slope (dashed line) and lagged cycle (dash-dotted line) loadings are modestly higher than in the nonswitching model. The cycle variance is half again as large as the trend variance in the low-variance state but triple the trend variance in the high-variance state.

Estimated term premia show large differences between high- and low-variance regimes at the short end of the yield curve. In contrast, the premium for longer maturities is neither economically nor statistically much different across regimes (see Figure 6). Note that the high-state trend variance is double that in the low state, while for the cycle, the high-state variance is quadruple that in the low state. The estimated premia are consistent with our model specification. Since the shorter yields load more on the cycle, the shorter yields have a larger difference in variance across regimes.

Trend and cycle estimates allowing for switching are shown in Figures 2a and 2b, using the regime-probability weighted estimates of the 120-month premium to make them comparable to cross-section estimates. The new estimates of trend and cycle differ little from the nonswitching estimates, which is the expected outcome given that the short-rate parameters do not switch and that we find little difference in the long-term premium.

Allowing for regime switching improves the fit of the model: the measurement errors variances are significantly smaller than those in the baseline model. The shorter yields, which are fitted rather badly in the baseline model, now have measurement errors with standard deviations of less than 30 basis points. Among the 17 yields, the 9-month yield has the worst fit, and the standard deviation for its measurement error is 28 basis points. Figure 5 compares the trend and cycle of model (adjusted by the term premium) with the level and slope factors estimated in cross sections.

2.4 Comparisons with Other Models

Both our model and that of Diebold and Li (2006) fit cross-sections of the term structure with the Nelson–Siegel curve. Diebold and Li evolve the cross-sections through time by fitting the three Nelson–Siegel factors to AR(1) processes. In contrast, our law of motion comes from the unobserved components model for the short rate, plus a contribution from Markov switching in the premium. Which model gives better out-of-sample forecasts? For short-horizon forecasts, Diebold–Li should have an advantage as they model both the short and long rate directly as AR(1) processes. However, if our model is correct, it should outperform Diebold–Li forecasts at longer horizons due to the inclusion of a nonstationary component. This turns out to be exactly the case.

Figure 7 shows 5 years of root mean square errors (RMSEs) for recursive out-of-sample forecasts for our model, the Diebold–Li model, as well as for a random walk model.[4] The top panel shows 1-month-ahead forecasts. All three models forecast well

4. We estimated the first two models through 1994:12 and then made out-of-sample forecasts out to a horizon of 3 years. We then moved ahead 1 month, reestimated and reforecast. Specifically, starting with

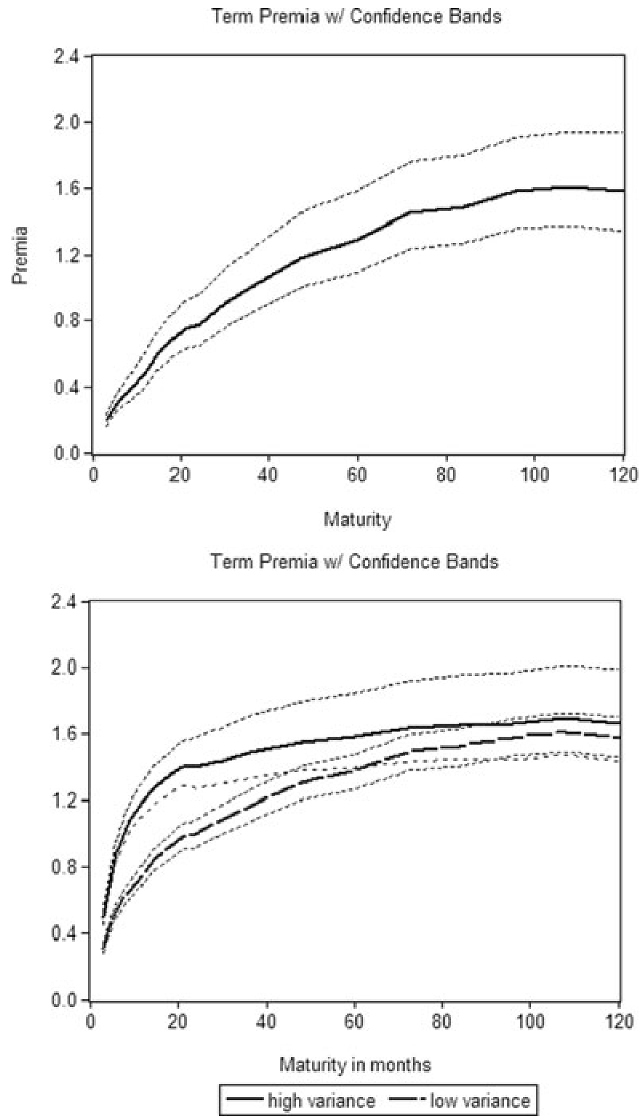


FIG. 6. Estimated Term Premia.

NOTES: The upper graph shows the estimated premium, plotted against the time to maturity, with the 95% bands. The lower graph shows the same estimates for the regime-switching model, separately for the two regimes.

across all maturities, the RMSE being on the order of 25–30 basis points. Diebold–Li outperforms our model by 5–10 basis points. The panels showing the RMSE 2 and 3

the smoothed state probability at time T, we forecast state probabilities out to the forecast horizon and then used a weighted average term premium. The cycle is forecast using the smoothed cycle estimate at time T and the estimated AR coefficients. The trend forecast is simply the time T trend estimate.

years out present a very different picture. For longer maturities and longer horizons our model outperforms Diebold–Li, the largest gaps being about 70 basis points. In summary, as one would expect from the differing stochastic specifications, our model is notably better at out-of-sample forecasting, primarily at longer horizons. As usual, a random walk model also performs quite well. As might be expected, for long horizon/long-maturity forecasts, our model and the random walk model look much alike.

While our core model focuses on expectations of future rates, the Markov-switching version introduces a time-varying risk premium. Does this help explain the failure of the expectations hypothesis as observed in the literature? The answer is both "no" and "yes."

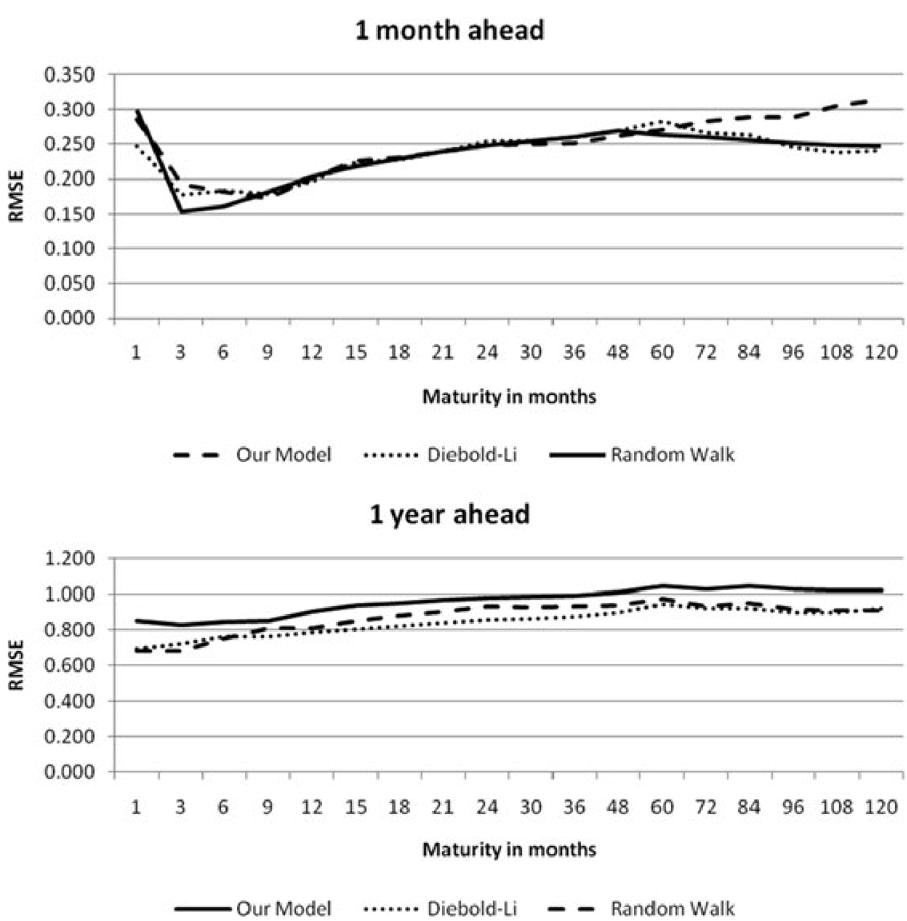


FIG. 7. Out-of-Sample Root Mean Square Errors (Our Model, Diebold–Li, and Random Walk).

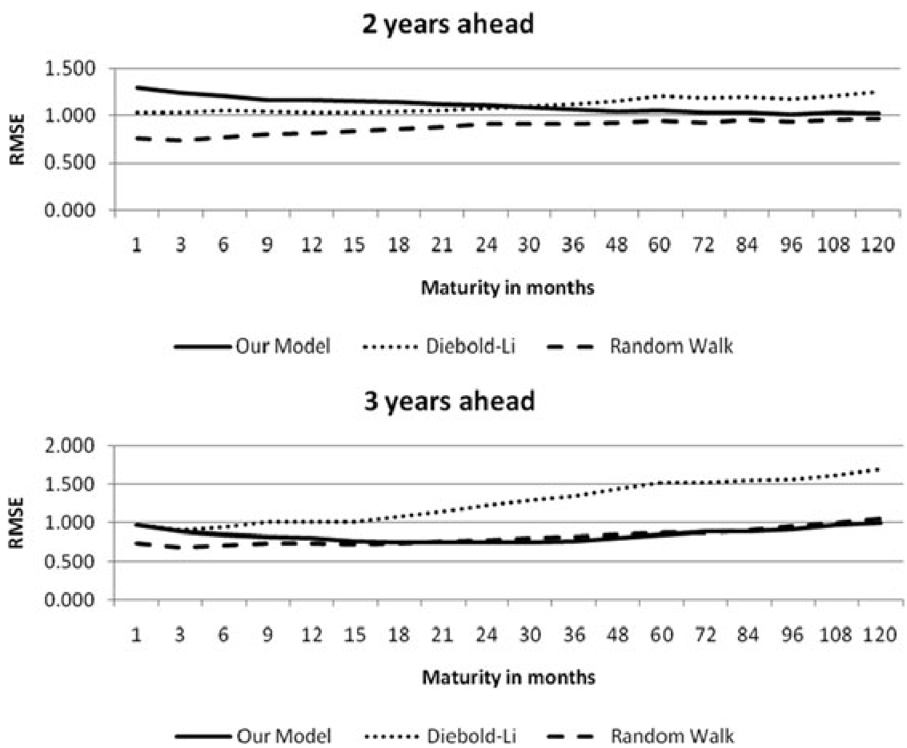


FIG. 7. Continued.

Consider the relationship between current and future m-period yields and the current $2m$-period yield

$$r_{t+m}^{m} - r_t^{2m} = \left(2\bar{\omega}^{(m)} - 2\bar{\omega}^{(2m)}\right) + \beta\left(r_t^{2m} - r_t^{m}\right) \\ + \left[\varepsilon_{t+m} + \left\{\left(\omega_{S_{t+m}}^{(m)} + \omega_{S_t}^{(m)} - 2\bar{\omega}^{(m)}\right) - \left(2\omega_{S_t}^{(2m)} - 2\bar{\omega}^{(2m)}\right)\right\}\right], \tag{20}$$

where ε_{t+m} is a forecast error and the constant term gives the difference in the unconditional means of the term premia. Equation (20) is what Thornton (2006) calls the "contrarian" specification of the Campbell and Shiller (1991) test. If the weak-form expectations hypothesis holds, that is, if the premium ω is time invariant, then a least squares regression of equation (20) should find $\beta = 1$. Empirically $\beta < 1$, which is evidence against the expectations hypothesis.

If we knew the states, we could augment the slope term, $r_t^{2m} - r_t^{m}$, in equation (20) with S_{t+m} and S_t. This would remove the state-dependent premia from the error term. Unfortunately we do not observe S_t, but only a noisy measure, $\Pr(S_t = 1)$, the smoothed probability. In Table 7, we show two different augmentations of equation (20). In the *ex ante* column, we include $\Pr(S_t = 1)$ as well as $r_t^{2m} - r_t^{m}$. This column

TABLE 7

CONTRARIAN CAMPBELL–SHILLER REGRESSIONS AUGMENTED WITH SMOOTHED STATE PROBABILITIES

	ex ante		*ex post*		
m	$r_t^{2m} - r_t^m$	$\Pr(S_t = 1)$	$r_t^{2m} - r_t^m$	$\Pr(S_t = 1)$	$\Pr(S_{t+m} = 1)$
3	−0.576	−0.160	−0.616	−0.399	0.538
	(0.368)	(0.208)	(0.341)	(0.190)	(0.161)
6	−0.606	−0.406	−0.312	−0.636	1.066
	(0.509)	(0.296)	(0.525)	(0.283)	(0.207)
9	−1.144	−0.461	−0.706	−0.541	1.425
	(0.354)	(0.363)	(0.326)	(0.349)	(0.269)
12	−1.016	−0.293	−0.486	−0.282	1.844
	(0.409)	(0.433)	(0.363)	(0.387)	(0.295)
15	−1.068	−0.407	−0.476	−0.290	2.220
	(0.617)	(0.452)	(0.597)	(0.352)	(0.316)
18	−0.954	−0.412	−0.495	−0.309	2.373
	(0.695)	(0.475)	(0.632)	(0.386)	(0.349)
24	−0.804	−0.171	−0.705	−0.355	2.345
	(0.727)	(0.528)	(0.662)	(0.440)	(0.413)
30	−0.468	−0.148	−0.301	−0.260	2.164
	(0.779)	(0.526)	(0.664)	(0.474)	(0.465)
36	0.213	0.071	0.056	−0.098	1.960
	(0.622)	(0.612)	(0.579)	(0.577)	(0.531)
48	0.232	−0.314	0.452	−0.377	2.214
	(0.692)	(0.691)	(0.627)	(0.584)	(0.639)
60	0.012	−0.739	0.759	−0.356	2.537
	(0.817)	(0.673)	(0.748)	(0.616)	(0.687)

(Newey–West standard errors in parentheses)

adds a minimal amount of information that would be available to an econometrician at time t, say, for pricing purposes, and should pick up time variation in the *difference* between long and short premia, $\omega_{S_t}^{(m)} - 2\omega_{S_t}^{(2m)}$.[5] Estimated β remains below 1, although β is significantly different from one at the 5% level only for maturities of 2 years or less. The smoothed probability is never significant even at the 10% level. Thus, from the point of view of the econometrician standing at time t, the answer to the question poised above is mostly "no."

In contrast, ask whether some part of the risk premium could be explained by a simple two-state model if we observed the states. In other words, might there be an *ex post* explanation? On the right side of Table 7, we include both of our noisy state measures, $\Pr(S_t = 1)$ and $\Pr(S_{t+m} = 1)$, the latter picking up the *level* of $\omega(m)_{St+m}$. We still find $\beta < 1$, although point estimates generally move closer to one. What we do find is that $\Pr(S_{t+m} = 1)$ is highly significant, with ts in the 4–7 range. Putting these observations together, a two-state model appears to explain a statistically significant, but quantitatively small, part of the time-varying risk premium. Thus, the answer is "yes" above—the two-state model helps explain the failure of the expectations hypothesis but does not resolve the issue.

5. In principle, one could forecast $\Pr(S_{t+m}= 1 \mid \Phi_t)$, but as a practical matter, the estimated Markov probabilities are such that this returns to the unconditional probability fairly quickly.

TABLE 8

Gibbs Sampling Results with Longer CRSP Data: Covariance Matrix for the State Variables, the AR(2) Parameters, and the Transitional Probabilities

	Posterior for baseline model				Posterior for model with switching			
	Mean	SD	95% bands		Mean	SD	95% bands	
Trend variance	0.213	0.019	0.187	0.245				
Cycle variance	0.522	0.053	0.448	0.617				
Covariance	−0.153	0.034	−0.213	−0.108				
Low trend variance					0.146	0.015	0.124	0.173
Low cycle variance					0.223	0.026	0.188	0.269
Low covariance					−0.109	0.020	−0.145	−0.083
High trend variance					0.519	0.130	0.354	0.744
High cycle variance					1.808	0.407	1.291	2.538
High covariance					−0.386	0.220	−0.775	−0.123
phi 1	0.684	0.043	0.614	0.754	0.696	0.038	0.632	0.758
phi 2	0.168	0.042	0.101	0.237	0.183	0.036	0.125	0.242
q (low var prob)					0.953	0.010	0.936	0.968
p (high var prob)					0.832	0.032	0.776	0.882

TABLE 9

GIBBS SAMPLING RESULTS WITH LONGER CRSP DATA: VARIANCE OF THE MEASUREMENT ERRORS

	Posterior for baseline model					Posterior for model with switching				
Maturity	Mean	SD	95% bands		R^2	Mean	SD	95% bands		R^2
3	0.133	0.008	0.120	0.147	0.983	0.087	0.005	0.078	0.096	0.989
6	0.262	0.016	0.237	0.289	0.967	0.174	0.010	0.158	0.192	0.978
12	0.135	0.009	0.121	0.151	0.983	0.074	0.005	0.067	0.082	0.990
24	0.036	0.002	0.032	0.040	0.995	0.023	0.001	0.021	0.025	0.997
36	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.001	1.000
48	0.027	0.002	0.025	0.030	0.996	0.020	0.001	0.018	0.022	0.997
60	0.069	0.004	0.062	0.077	0.989	0.049	0.003	0.044	0.055	0.993

TABLE 10

GIBBS SAMPLING RESULTS WITH LONGER CRSP DATA: TERM PREMIA

	Posterior for baseline model				Posterior for model with switching							
					Low variance regime				High variance regime			
Maturity	Mean	SD	95% bands		Mean	SD	95% bands		Mean	SD	95% bands	
3	0.364	0.015	0.339	0.389	0.251	0.014	0.229	0.274	0.817	0.028	0.771	0.863
6	0.553	0.022	0.517	0.589	0.400	0.022	0.363	0.436	1.185	0.042	1.116	1.254
12	0.745	0.019	0.715	0.776	0.616	0.024	0.568	0.649	1.323	0.042	1.261	1.400
24	0.925	0.018	0.900	0.957	0.857	0.031	0.788	0.898	1.303	0.054	1.230	1.408
36	1.088	0.018	1.067	1.120	1.072	0.036	0.989	1.123	1.278	0.062	1.198	1.404
48	1.217	0.020	1.190	1.254	1.242	0.040	1.151	1.295	1.260	0.068	1.172	1.397
60	1.293	0.023	1.261	1.333	1.343	0.043	1.247	1.399	1.247	0.073	1.149	1.391

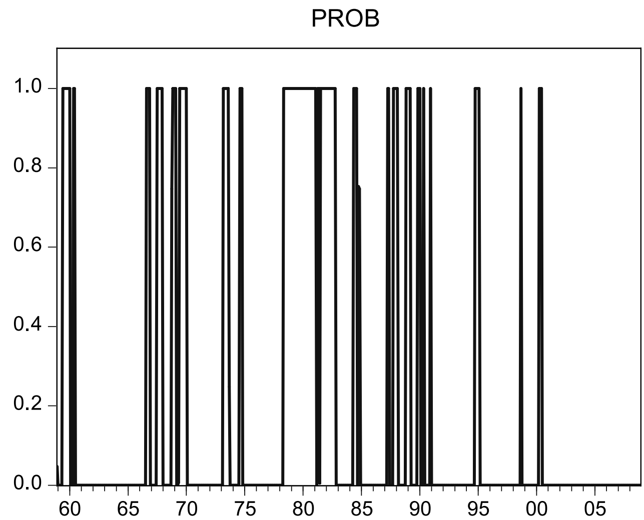


FIG. 8. Probability of High-Variance Regime for the Longer CRSP Data.

3. RESULTS FROM ANOTHER DATA SET

To make sure that our results are not sensitive to the sample period and the choice of maturities, we reestimate our model using the 1-month, 3-month, 6-month, 1-year, 2-year, 3-year, 4-year, and 5-year zero-coupon yields from the CRSP data set, over the period from December 1958 to December 2008.[6] The longer sample period allows us to conduct a more reliable out-sample forecasting exercise, albeit at the cost of foregoing maturities longer than 5 years. Posterior results for the parameters based on the whole sample are shown in Tables 8–10. The persistence of the cyclical component, relative size of the two shocks, and the goodness of fit are all similar to the results with the Bliss data. The only difference between the two sets of results is that both the high-variance and low-variance regimes are more persistent. Figure 8 plots the posterior probability of being in the high-variance regime, and the classification of the regimes is similar to that with the Bliss data in Figure 5.

One problem with conducting out-of-sample forecasting with the Bliss data is that few observations are available for the out-of-sample forecast. In addition, the 1996–2000 period was fairly quiet in terms of interest rate fluctuations, with the benchmark 3-month Treasury bill moving across a 221 basis point range. With 50 years of CRSP data, we can leave the last 10 years of data for out-of-sample forecast of 1-month, 1-year, 2-year, 3-year, and 4-year horizons. That is, we first estimate our model using data up to December 1997, make forecasts, and then reestimate the

6. Source: © 201002CRSP®, Center for Research in Security Prices. Booth School of Business, The University of Chicago. Used with permission. All rights reserved. www.crsp.chicagobooth.edu.

model by adding the January 1998 data. This is also a more interesting period over which to forecast, as the 3-month Treasury bill moved across a 614 basis point range.

In Figure 9, we plot the RMSE of our model, the Diebold–Li model and a simple random walk for the eight yields. In contrast to the forecasting results using the Bliss data, the Diebold–Li model's RMSE is higher than a simple random walk for almost

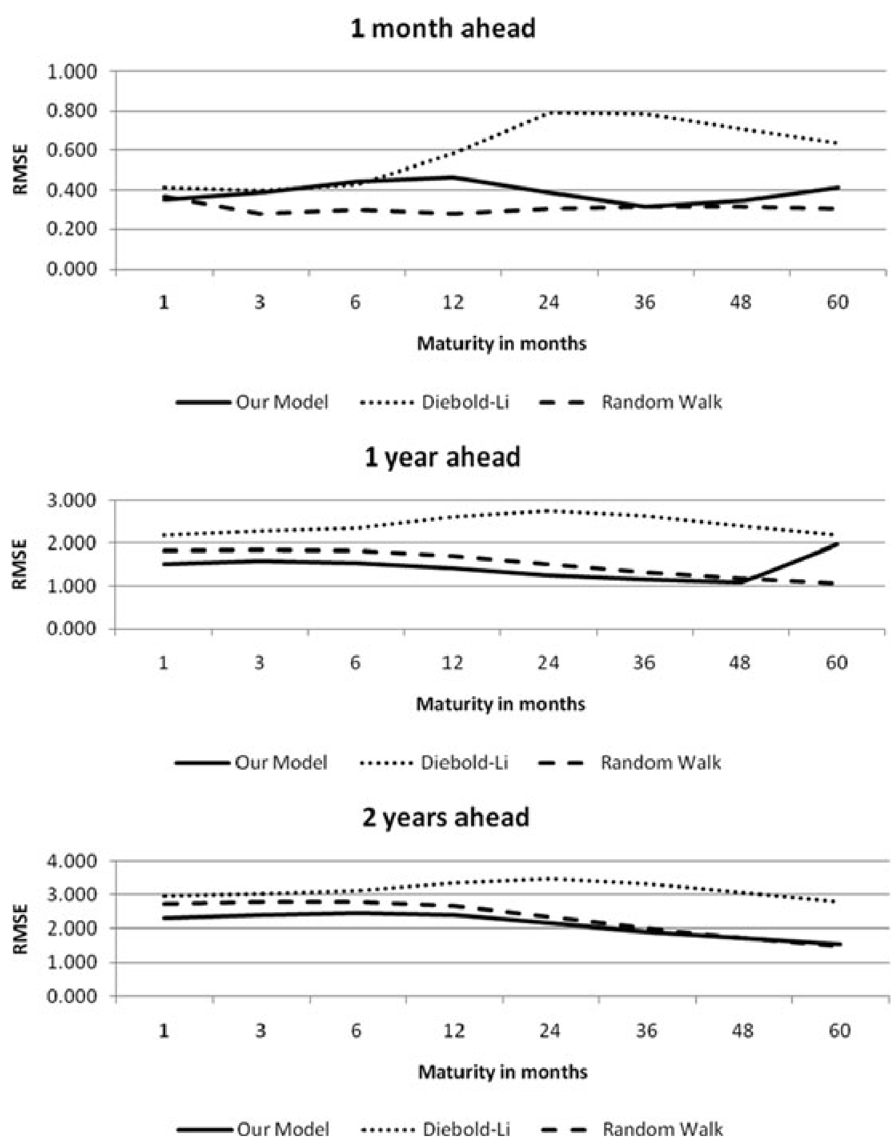
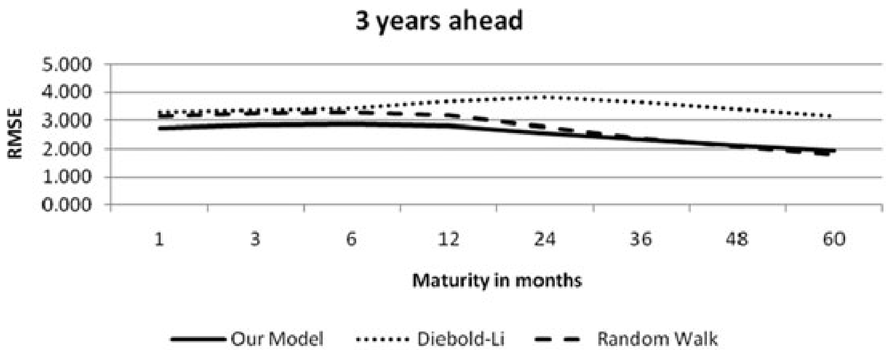


FIG. 9. Out-of-Sample Root Mean Square Errors with Longer CRSP Data (Our Model, Diebold–Li, and Random Walk).



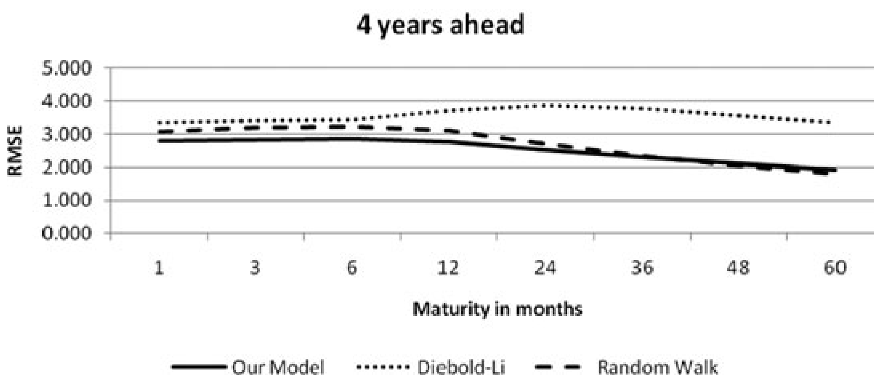


FIG. 9. Continued.

all horizons and maturities. Except the 1-month horizon, our model outperforms the random walk in most cases.[7]

4. CONCLUDING REMARKS

Taking a trend/cycle unobserved components time-series model of the short rate and projecting the short rate into the future to give an expectations hypothesis with constant term premium model of the yield curve works extremely well in the sense of giving consistent time series/cross-section fits of yields that fit the data well. We identify two clear factors, trend and cycle, and show that the third factor allowed by the model does not seem to be very important. Model-based measures of uncertainty do an intriguingly good job of predicting estimated term premia. Allowing for regime switching in shock variances improves model performance yet more, giving more definite identification of an unimportant third factor. Regime switching also introduces time varying term premia into the model in a very natural way. Finally, while

7. Thanks to an anonymous referee for suggesting we look at more recent data.

we have not pursued it in this paper, our model may prove useful in integrating the term structure into models of broader macroeconomic behavior.

LITERATURE CITED

Bliss, Robert R. (1997) "Testing Term Structure Estimation Methods." In *Advances in Futures and Options Research,* edited by P. Boyle, G. Pennacchi, and P. Ritchken, pp. 197–231. Greenwich, CT: JAI Press.

Campbell, John Y., and Robert J. Shiller. (1991) "Yield Spreads and Interest Rate Movements: A Birds Eye View." *Review of Economic Studies,* 58, 495–514.

Cochrane, John H. (1988) "How Big Is the Random Walk in GNP?" *Journal of Political Economy,* 96, 893–920.

Christenson, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch. (2009) "An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model." *Econometrics Journal,* 12, 33–64.

De Jong, Frank. (2000) "Time Series and Cross-Section Information in Affine Term-Structure Models." *Journal of Business & Economic Statistics,* 18, 300–14.

Diebold, Francis X., and Canlin Li. (2006) "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics,* 130, 337–64.

Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba. (2006) "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach." *Journal of Econometrics,* 131, 309–38.

Duffee, Gregory R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance,* 57, 405–43.

Fama, Eugene, and Robert R. Bliss. (1987) "The Information in Long-Maturity Forward Rates." *American Economic Review,* 77, 680–92.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. (2007) "The U.S. Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics,* 54, 2291–2304.

Nelson, Charles R., and Andrew F. Siegel. (1987) "Parsimonious Modeling of Yield Curves." *Journal of Business,* 60, 473–89.

Startz, Richard. (1982) "Do Forecast Errors or Term Premia Really Make the Difference between Long and Short Rates?" *Journal of Financial Economics,* 10, 323–29.

Thornton, Daniel L. (2006) "Tests of the Expectations Hypothesis: Resolving the Campbell-Shiller Paradox." *Journal of Money, Credit and Banking,* 38, 511–41.