The Yield Curve Through Time and Across Maturities [*]

Richard Startz and Kwok Ping Tsang

This Draft: 3/8/2007

Abstract: 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. 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 that in the low-variance regime. The estimation results support our model as a simple and yet reliable framework for modeling the term structure.

JEL Classifications: C22, E32, E43

Keywords: term structure of interest rates, Nelson-Siegel yield curve, trend-cycle decomposition.

[*] Startz: Department of Economics, University of Washington, Box 353330, Seattle, WA 98195 (startz@u.washington.edu); Tsang: Department of Economics, University of Washington, Box 353330, Seattle, WA 98195 (byront@u.washington.edu). Financial support from the Castor Professorship at the University of Washington is gratefully acknowledged. The authors thank Charles R. Nelson for helpful comments, and Robert Bliss for making his interest rate data available.

I. Introduction

Some two decades ago, Charles Nelson and Andrew 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 zero 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 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 long rates 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 theory and the estimated results.[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 an expectations-hypothesis model of the term structure with zero term premia. Zero term premia is empirically untenable (Startz (1982)). 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.

We then 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

[1] For a theoretically consistent approach tying together time series and cross-section in affine models, see de Jong (2000).

process. Because modeling the cross sections adds very few parameters to the time-series model, a great deal of data is 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 easier to relate to other macro decompositions, notably GDP. We compare our estimates of trend and cycle to appropriate level and slope estimates 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 isn't 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, we find higher premia in higher variance states.

II. A Serious, Stochastic Nelson-Siegel Model

A. The Original Nelson-Siegel Model

The original setup Nelson and Siegel article,. the process for the short rate r (with maturity of one period) is assumed to be a non-homogeneous second-order differential equation,

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

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

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

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

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

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

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

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

In practice, equation (3) is estimated by OLS period by period, fixing the value of $1/\kappa$ at 0.0609. Any restrictions which 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 which summarize the dynamics of the term structure over time. The flexible functional form of (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.

B. 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,

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

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

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

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

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

$$r_t = r_t^{(1)} = \tau_t + c_t,$$

$$r_t^{(2)} = \omega^{(2)} + \frac{1}{2}\left(r_t + E_t r_{t+1}\right) = \omega^{(2)} + \tau_t + \frac{1+\phi_1}{2} c_t + \frac{\phi_2}{2} c_{t-1}, \tag{9}$$

$$r_t^{(3)} = \omega^{(3)} + \frac{1}{3}\left(r_t + E_t r_{t+1} + E_t r_{t+2}\right) = \omega^{(3)} + \tau_t + \frac{1+\phi_1+\phi_1^{\,2}+\phi_2}{3} c_t + \frac{\phi_2+\phi_1\phi_2}{3} c_{t-1},$$

and so on. Note In general we can write:

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

where the functions $f(\bullet)$ and $g(\bullet)$ are the factor loadings. As we show in Appendix A, 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 (12).

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

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

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 were 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 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 zero and the factor loading on the slope equals $1/m$. The yield curve would at all times be a hyperbola with a half-life of two months. Neither horizontal line nor short-lived 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 five basis points occurs less than five percent 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. Additionally, 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 representations 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 five percent 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 which incorporates both times series and cross section information. Let the *M*-vector of interest rates at time *t* be $\vec{r}_t = \left[r_t^{(1)} \; r_t^{(3)} \; ... \; r_t^{(120)} \right]'$. We augment the time series model for the short rate given in equations (5)-(8) with the yield curves in equation (13).

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

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 $f(\bullet)$ and $g(\bullet)$ are functions of the two AR(2) coefficients.

Our state space model imposes 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 4 ARMA parameters plus M term premia. (A second model, below, adds another $M + 2$ parameters.) Note that the 3-latent factor model of Diebold, Rudebusch and Aruoba (2006) effectively uses only 3 means and 3^2 VAR(1) parameters.

III. Empirical Estimates

A. 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 ten-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 $V_k = \mathrm{var}\left(r_{t+k}^{(m)} - r_t^{(m)}\right) / k$, 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 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.

B. Estimation of the baseline model

We estimate the model (5)-(8) and (13) by Gibbs sampling organized in two steps. Given the parameters $\left\{\Sigma, \Omega, \sigma_u^2, \sigma_v^2, \sigma_{uv}, \phi_1, \phi_2\right\}$ 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 appear in Appendix B.

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 UC models there is always an identification issue in ensuring that the estimated cycle is stationary, we use persistent but

---

[2] The dataset 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.

stationary priors for the AR coefficients, $N\left(\begin{bmatrix}1.1\\-.2\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix}\right)$. 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 percent confidence bands appear in Tables 4a-c. 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 percent 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 percent band $(-0.050, 0.134)$. Estimates of term premia rise with maturity, are essentially identical—within three basis points of—mean differences in yields, and are tightly estimated with standard deviations between two and eight basis points.

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.

Our model performs well, although not perfectly, in fitting the yields across time and maturity. 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

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

The conditional variance is a function of the model parameters, $\operatorname{var}\left(r^{(m)}\right) = \sigma_u^2 + f^2\left(m\right)\sigma_v^2 + 2f\left(m\right)\sigma_{uv}$. Since the price of a bond is $p^{(m)} = 1\big/\left(1+r^{(m)}\right)^m$, the usual Taylor series approximation gives $\operatorname{var}\left(p^{(m)}\right) \approx m^2 p^2 \operatorname{var}\left(r^{(m)}\right)$. We compute this variance using the estimated parameters and mean prices[4]. Figure 4 shows the relation between estimated term premia and uncertainty. Visually quite strong, the correlation across maturities between the premium and standard deviation is 0.97. 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 5 basis point increase in the premium.

C. 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

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

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

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

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

---

[4] Using mean prices in the approximation adds an element of endogeneity, but note that a high premium in the yield gives a lower mean price and that we find a positive relationship between computed uncertainty and premium.

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

$$\Pr\left[S_t = 0 \mid S_{t-1} = 0\right] = q \text{ and } \Pr\left[S_t = 1 \mid S_{t-1} = 1\right] = p \qquad (19)$$

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 (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 non-switching 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, premium for longer maturities are neither economically nor statistically important 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 non-switching 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 seventeen yields, the nine-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.

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

References

1) **Bliss, Robert R.** (1997), "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, Vol. 9.
2) **Cochrane, J. H.** (1988), "How Big is the Random Walk in GNP?" *Journal of Political Economy*, 96, 5, 893-920.
3) **De Jong, Frank**. (2000), "Time Series and Cross-Section Information in Affine Term-Structure Models," *Journal of Business & Economic Statistics*, 18, 3, 300-314.
4) **Diebold, F.X. and Li., C.** (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337-364.
5) **Diebold, F.X., Rudebusch, G.D. and Aruoba, S.B.** (2006), "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics*, 131, 309-338.
6) **Diebold, F.X., Piazzesi, M., and Rudebusch, G.D.** (2005), "Modeling Bond Yields in Finance and Macroeconomics, The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *American Economic Review*, 95, 415-420
7) **Duffee, G.** (2002) "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405–443.
8) **Fama, E. and Bliss, R.** (1987), "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680-692.
9) **Granger, C.W.J. and Newbold, P**. (1986), *Forecasting Economic Time Series*. Academic Press, New York.
10) **Gürkaynak, R.S., Sack, B. and Wright, J.H.** (2006), "The U.S. Treasury yield curve: 1961 to the present,", *Finance and Economics Discussion Series* 2006-28, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.
11) **Kim, C.-J. and Nelson, C.R.** (1998), *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge: MIT Press.
12) **Nelson, C.R. and Siegel, A.F.** (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 473-489.
13) **Startz, R.** (1982), "Do Forecast Errors or Term Premia Really Make the Difference Between Long and Short Rates?" *Journal of Financial Economics*, l0 (3), pp. 323-329.

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 (0.486)	0.860 (0.373)
ϕ_2				-0.217 (0.314)
θ_1		0.072 (0.052)	0.515 (0.464)	-0.807 (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

Table 2 – Descriptive Statistics for the Yield Data

Maturity	Mean	Median	Maximum	Minimum	S.D.
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

Table 3 – Cochrane Variance Ratio for Six Yields

	Maturity					
Horizon in months	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

?

Table 4a – 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 4b – Gibbs Sampling Results: Variance of the Measurement Errors

	Posterior for Baseline Model					Posterior for Model with Switching				
Maturity	Mean	SD	95% bands		R-sq	Mean	SD	95% bands		R-sq
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 4c – 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

Figure 1 – One Month Treasury Bill Yields

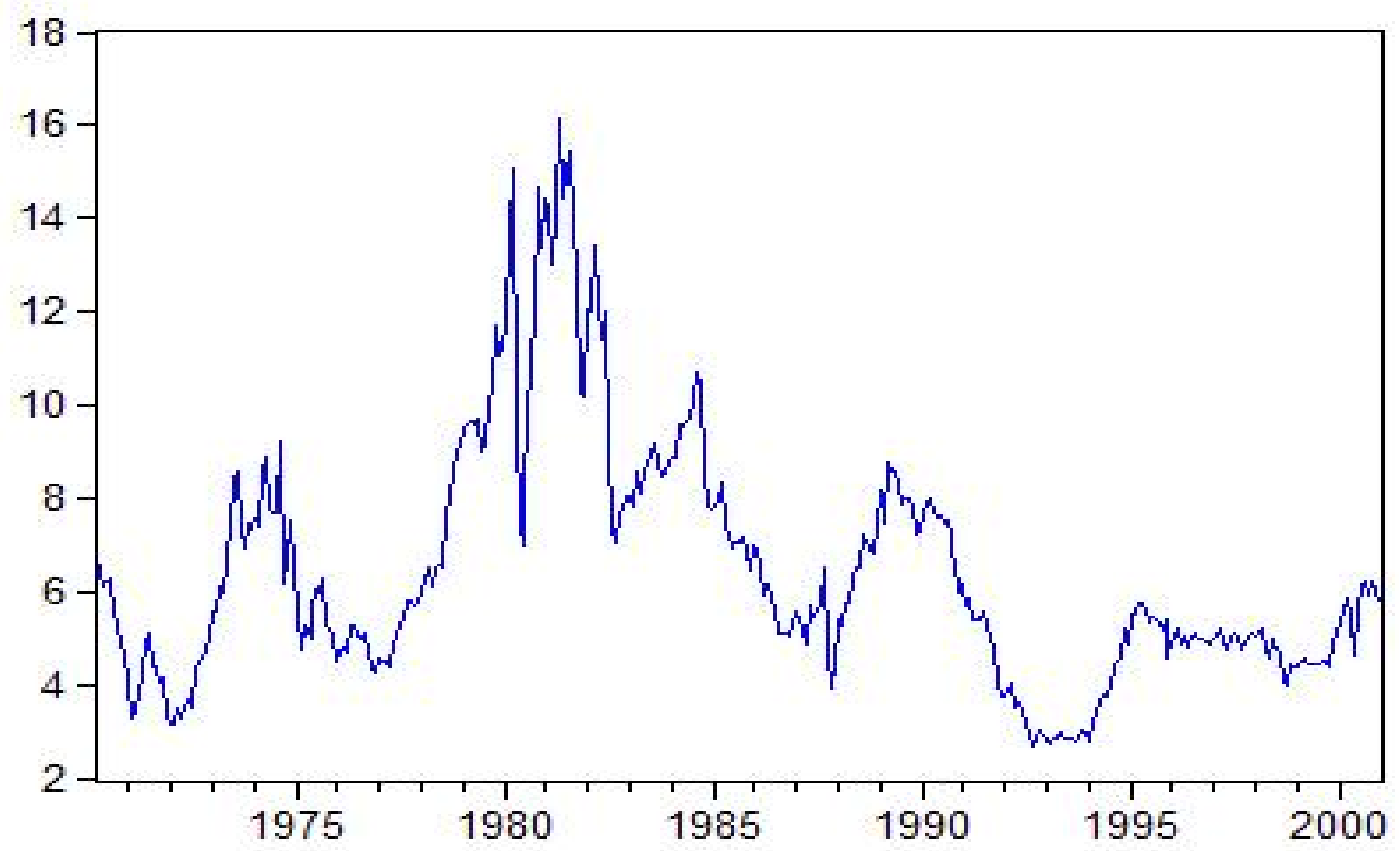


Figure 2a – Estimated Trend Versus Estimated Cross-Section Level Factors

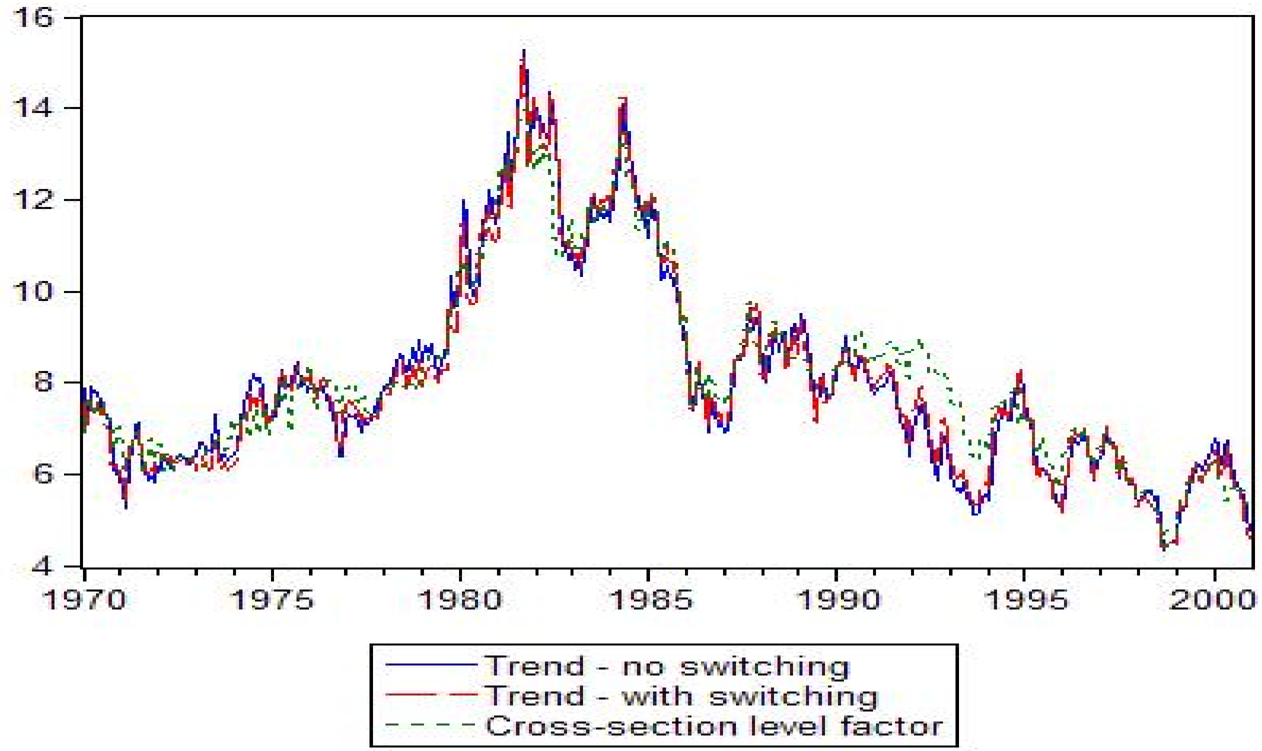


Figure 2b – Estimated Cycle Versus Estimated Cross-Section Slope Factors

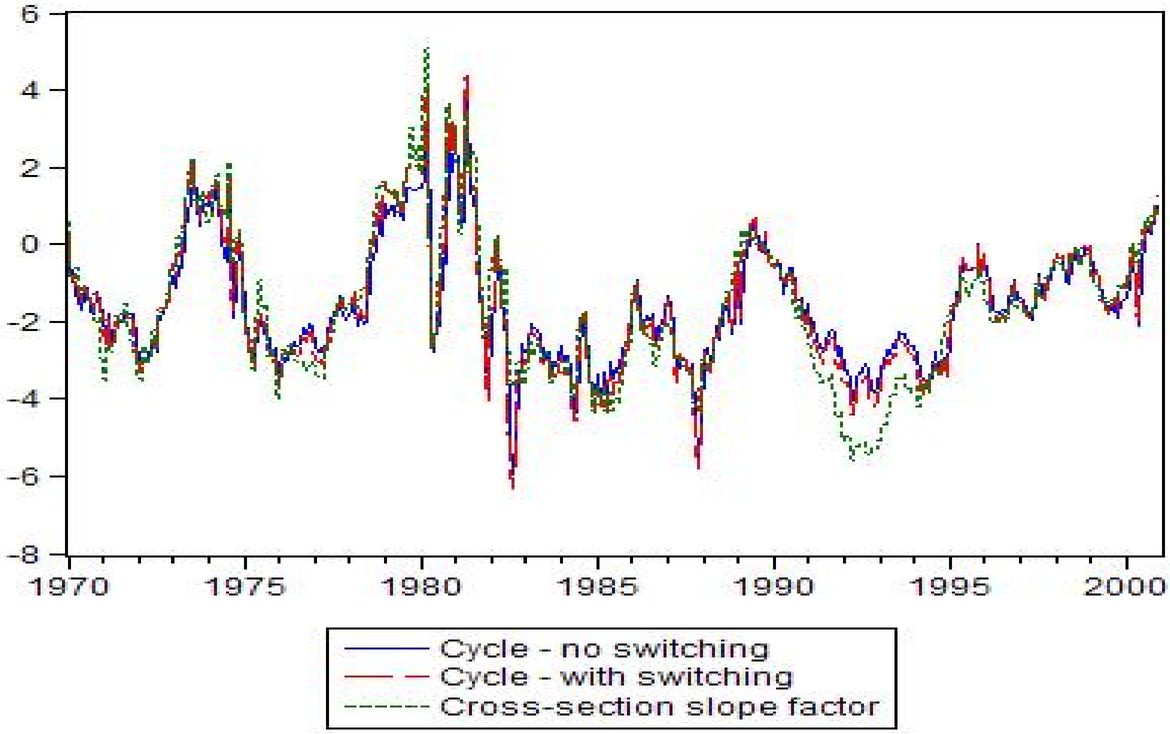


Figure 3 – Slope and Lagged Cycle Loadings

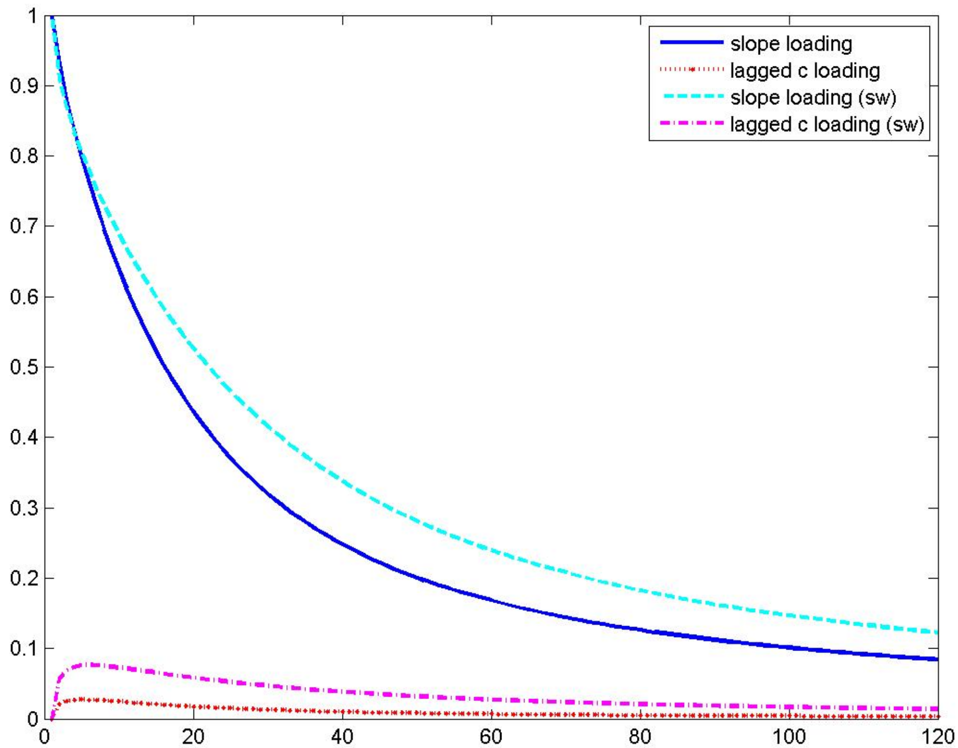


Figure 4- Term Premia versus Conditional Standard Deviation of Prices.

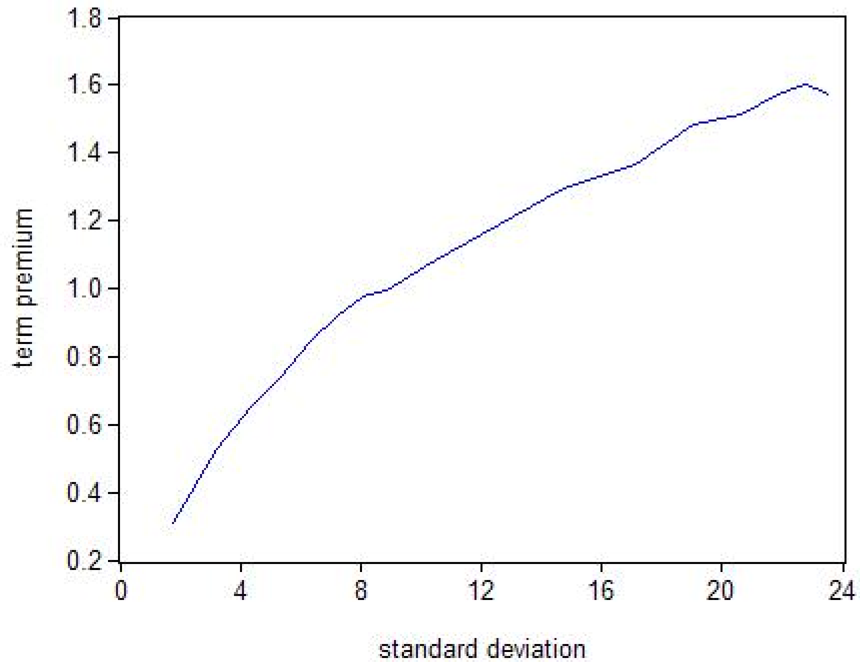


Figure 5 – Probability of High-Variance Regime

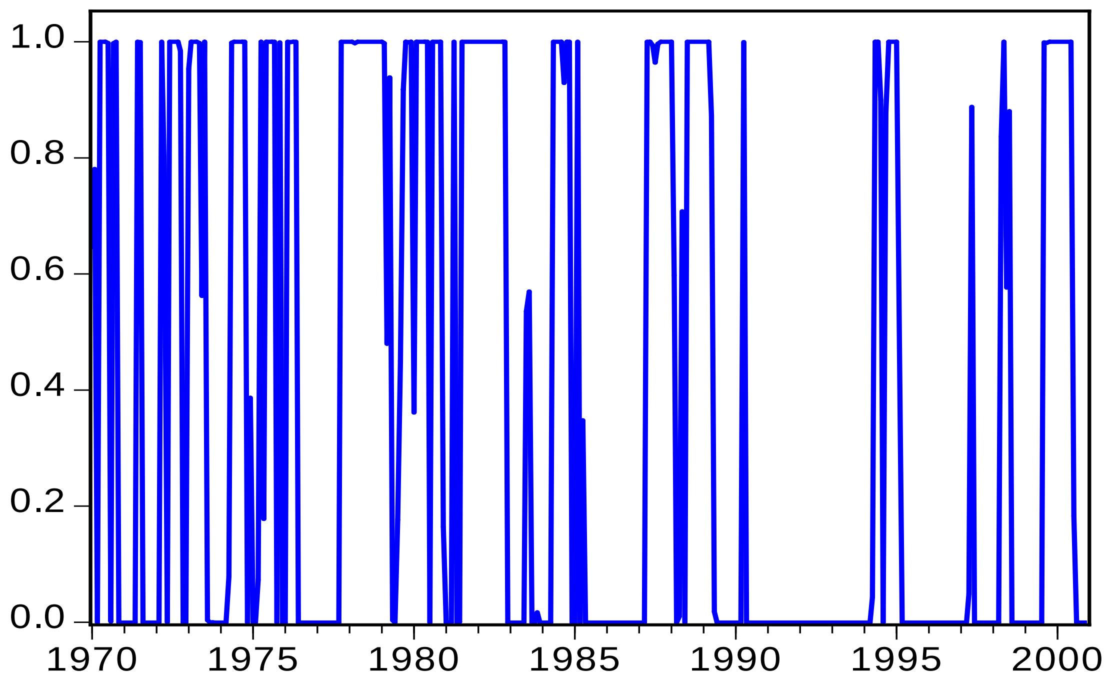
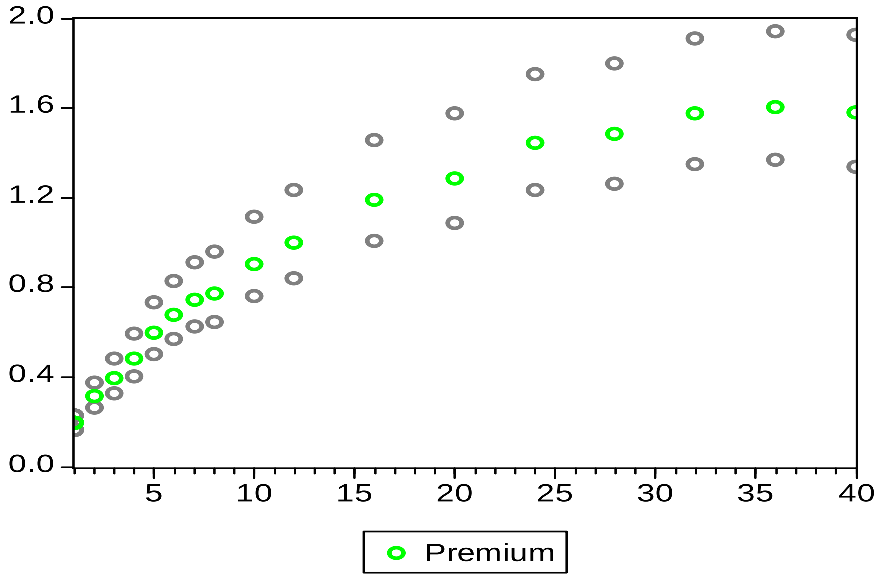
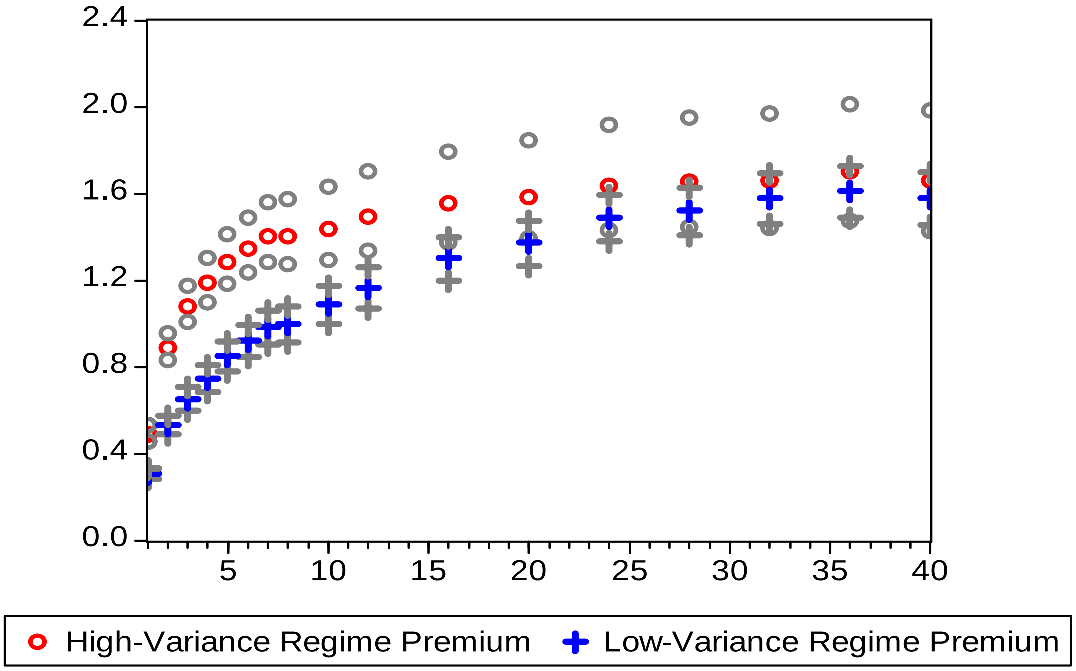


Figure 6 – Estimated Term Premia





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.

Appendix A – Solving the Discrete-Time Yield Curve Not for Publication

Consider the cycle part c_t as a second-order difference equation $c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2}$ (the stochastic part is left out since we are taking expectation). We aim at solving for c_{t+k} as a function of only c_t and c_{t-1} (our factors). To make the timing convention less confusing, we rename time t as 0, and define c_0 and c_{-1} to be the initial conditions. Our AR(2) setup does not exclude unequal roots in the characteristic equation, but to be consistent with the NS setup, we assume equal real roots $\eta = \phi_1 / 2$ for now (which is equal to the restriction $\phi_2 = -\phi_1 / 4$). The general solution is $c_k = P_1 \eta^k + P_2 k \eta^k$.

To determine the constants, use the initial conditions $c_0 = P_1$ and $c_{-1} = c_0 \eta^{-1} - P_2 \eta^{-1}$, then we have $c_k = c_0 \eta^k + \left(c_0 - \eta c_{-1}\right) k \eta^k$. "Shifting" the time by t (which does not affect the result) we have $c_{t+k} = c_t \eta^k + \left(c_t - \eta_{t-1}\right) k \eta^k$. Summing up it becomes to $\frac{1}{k}\sum_{i=0}^{k-1} c_{t+i} = \frac{1}{k}\left[c_t \sum_{i=0}^{k-1} \eta^i + \left(c_t - \eta c_{t-1}\right) \sum_{i=0}^{k-1} i\eta^i \right]$. By geometric sum we have $\sum_{i=0}^{k-1} \eta^i = \frac{1-\eta^k}{1-\eta}$. The other sum $\sum_{i=0}^{k-1} i\eta^i$ is solved as follows:

$$\sum_{i=0}^{k-1} i\eta^i = \eta + 2\eta^2 + 3\eta^3 + ... + \left(k-1\right)\eta^{k-1}$$

$$\left(1-\eta\right)\sum_{i=0}^{k-1} i\eta^i = \eta + 2\eta^2 + 3\eta^3 + ... + \left(k-1\right)\eta^{k-1} - \eta^2 - 2\eta^3 - 3\eta^4 ... - \left(k-2\right)\eta^{k-1} - \left(k-1\right)\eta^k$$

$$\left(1-\eta\right)\sum_{i=0}^{k-1} i\eta^i = \eta + \eta^2 + \eta^3 + ... + \eta^{k-1} - \left(k-1\right)\eta^k$$

$$\left(1-\eta\right)^2 \sum_{i=0}^{k-1} i\eta^i = \eta + \eta^2 + \eta^3 + ... + \eta^{k-1} - \left(k-1\right)\eta^k - \eta^2 - \eta^3 - ... - \eta^k + \left(k-1\right)\eta^{k+1}$$

$$\sum_{i=0}^{k-1} i\eta^i = \frac{\eta - k\eta^k + \left(k-1\right)\eta^{k+1}}{\left(1-\eta\right)^2}$$

So finally we have result [9]:

$$r_t^{(m)} = \tau_t + c_t \frac{1}{m}\left[\frac{1-\eta^m}{1-\eta} + \frac{\eta - m\eta^m + \left(m-1\right)\eta^{m+1}}{\left(1-\eta\right)^2} \right] - \eta c_{t-1} \frac{1}{m}\left[\frac{\eta - m\eta^m + \left(m-1\right)\eta^{m+1}}{\left(1-\eta\right)^2} \right]$$

We derive the case of unequal real roots [9'] in the same manner, using the general solution $c_k = P_1 \eta_1{}^k + P_2 \eta_2{}^k$ with the roots η_1 and η_2.

Appendix B - Details of the Gibbs Sampling Procedure – Not for Publication

This appendix describes the Gibbs sampling procedure. For further discussion please consult Kim and Nelson (1999), from which this appendix borrows.

First write our model in state-space form:

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

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

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

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

Rewrite the model in matrix notation by defining $\boldsymbol{\beta}_t = [\tau_t, c_t, c_{t-1}]'$:

$$\mathbf{r}_t = \omega_{S_t} + \mathbf{T}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\beta}_t = \mathbf{H}\boldsymbol{\beta}_{t-1} + \mathbf{n}_t,\ \mathbf{n}_t = [u_t \quad v_t \quad 0]'\ \mathbf{n}_t \sim N\left(\mathbf{0},\mathbf{Q}_0(1-S_t)+\mathbf{Q}_1 S_t\right),$$

where $\mathbf{Q}_0 = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{uv0} & 0 \\ \sigma_{uv0} & \sigma_{v0}^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\mathbf{Q}_1 = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{uv1} & 0 \\ \sigma_{uv1} & \sigma_{v1}^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Step 1: Given the data $\tilde{\mathbf{r}}_T = [\mathbf{r}_1\ \mathbf{r}_2\ \ldots\ \mathbf{r}_T]'$ and the starting values for the hyper-parameters $\{\boldsymbol{\Sigma},\omega_0,\omega_1,\sigma_{u0}^2,\sigma_{u1}^2,\sigma_{v0}^2,\sigma_{v1}^2,\sigma_{uv0},\sigma_{uv1},\phi_1,\phi_2,p,q\}$ and the state $\tilde{S}_T = [S_1\ S_2\ \ldots\ S_T]'$ (which is generated from the starting values for the transitional probabilities $\{p,q\}$), we first run the Kalman filter:

<u>Prediction</u>

1. $\boldsymbol{\beta}_{t|t-1} = \mathbf{H}\boldsymbol{\beta}_{t-1|t-1}$

2. $\mathbf{P}_{t|t-1} = \mathbf{H}\mathbf{P}_{t-1|t-1}\mathbf{H}' + \mathbf{Q}_{S_t}$

<u>Updating</u>

3. $\boldsymbol{\beta}_{t|t} = \boldsymbol{\beta}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{T}'\left(\mathbf{T}\mathbf{P}_{t|t-1}\mathbf{T}'+\boldsymbol{\Sigma}\right)^{-1}\left(\mathbf{r}_t - \mathbf{T}\boldsymbol{\beta}_{t|t-1} - \boldsymbol{\theta}_{S_t}\right)$

4. $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{T}'\left(\mathbf{T}\mathbf{P}_{t|t-1}\mathbf{T}'+\boldsymbol{\Sigma}\right)^{-1}\mathbf{T}\mathbf{P}_{t|t-1}$

At the end we obtain $\boldsymbol{\beta}_{t|t} = E(\boldsymbol{\beta}_t \mid \tilde{\mathbf{r}}_t)$ and $\mathbf{P}_{t|t} = Cov(\boldsymbol{\beta}_t \mid \tilde{\mathbf{r}}_t)$ for $t = 1, 2, ..., T$. Using $\boldsymbol{\beta}_{T|T}$ and $\mathbf{P}_{T|T}$, which contain information from the whole sample $\tilde{\mathbf{r}}_T$, we generate the last set of state variables based on the conditional distribution:

$\boldsymbol{\beta}_T \mid \tilde{\mathbf{r}}_T \sim N(\boldsymbol{\beta}_{T|T}, \mathbf{P}_{T|T})$,

and for the rest of the state variables, we use the conditional distribution:

$\boldsymbol{\beta}_t \mid \tilde{\mathbf{r}}_t, \boldsymbol{\beta}^*_{t+1} \sim N\left(\boldsymbol{\beta}_{t|t,\boldsymbol{\beta}^*_{t+1}}, \mathbf{P}_{t|t,\boldsymbol{\beta}^*_{t+1}}\right)$ for $t = T-1, T-2, ..., 1$

To obtain the two arguments in the conditional distribution, we first define $\boldsymbol{\beta}^*_{t+1} = \mathbf{H}^*\boldsymbol{\beta}_t + \mathbf{n}^*_{t+1}$, where $\boldsymbol{\beta}^*_{t+1}$ is the first two elements of $\boldsymbol{\beta}_{t+1}$ $\mathbf{T}^*$ is the first two rows of $\mathbf{T}$ and $\mathbf{n}^*_t$ is the first two elements of $\mathbf{n}_t$. Likewise we have $\mathbf{Q}^*_{S_t}$ as the first 2 by 2 block of $\mathbf{Q}_{S_t}$. The two arguments in the conditional distribution are given as:

$$\boldsymbol{\beta}_{t|t,\boldsymbol{\beta}^*_{t+1}} = E(\boldsymbol{\beta}_t \mid \tilde{\mathbf{r}}_t, \boldsymbol{\beta}^*_{t+1}) = \boldsymbol{\beta}_{t|t} + \mathbf{P}_{t|t}\mathbf{H}^{*}{}'\left(\mathbf{H}^*\mathbf{P}_{t|t}\mathbf{H}^{*}{}' + \mathbf{Q}^*_{S_{t+1}}\right)^{-1}\left(\boldsymbol{\beta}^*_{t+1} - \mathbf{H}^*\boldsymbol{\beta}_{t|t}\right)$$

$$\mathbf{P}_{t|t,\boldsymbol{\beta}^*_{t+1}} = Cov(\boldsymbol{\beta}_t \mid \tilde{\mathbf{r}}_t, \boldsymbol{\beta}^*_{t+1}) = \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{H}^{*}{}'\left(\mathbf{H}^*\mathbf{P}_{t|t}\mathbf{H}^{*}{}' + \mathbf{Q}^*_{S_{t+1}}\right)^{-1}\mathbf{H}^*\mathbf{P}_{t|t}$$

The above procedures generate the state variables $\tilde{\boldsymbol{\beta}}_T = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_T]'$, though we only keep the first two elements of each for future inference, i.e. $\tilde{\tau}_T = [\tau_1\ \tau_2\ ...\ \tau_T]'$ and $\tilde{c}_T = [c_1\ c_2\ ...\ c_T]'$.

Step 2: Given the hyper-parameters $\{\boldsymbol{\Sigma}, \omega_0, \omega_1, \sigma^2_{u0}, \sigma^2_{u1}, \sigma^2_{v0}, \sigma^2_{v1}, \sigma_{uv0}, \sigma_{uv1}, \phi_1, \phi_2, p, q\}$, the state variables $\tilde{\boldsymbol{\beta}}_T = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_T]'$ from Step 1, and the data $\tilde{\mathbf{r}}_T$, we first run the Hamilton filter:

1. $\Pr(S_t, S_{t-1} \mid \psi_{t-1}) = \Pr(S_t \mid S_{t-1})\Pr(S_{t-1} \mid \psi_{t-1})$
2. $f(\mathbf{r}_t \mid \psi_{t-1}) = \sum_{S_t}\sum_{S_{t-1}} f(\mathbf{r}_t \mid S_t, S_{t-1}, \psi_{t-1})\Pr(S_t, S_{t-1} \mid \psi_{t-1})$
3. $\Pr(S_t \mid \psi_t) = \sum_{S_{t-1}} \dfrac{f(\mathbf{r}_t \mid S_t, S_{t-1}, \psi_{t-1})\Pr(S_t, S_{t-1} \mid \psi_{t-1})}{f(\mathbf{r}_t \mid \psi_{t-1})}$

where $f(\bullet)$ is the normal density function, and ψ_{t-1} is the information available at $t-1$. The initial probability $\Pr(S_0 \mid \psi_0)$ is steady state probability determined by $\{p,q\}$. After the Hamilton filter, we obtain $\Pr(S_T \mid \psi_T)$ from which we generate S_T. For the state at time $t = T-1, T-2, ..., 1$, we use the iteration:

$$\Pr(S_t = 1 \mid S_{t+1}, \tilde{\mathbf{r}}_t) = \frac{\Pr(S_{t+1} \mid S_t = 1)\Pr(S_t = 1 \mid \tilde{\mathbf{r}}_t)}{\sum_{j=0}^{1} \Pr(S_{t+1} \mid S_t = j)\Pr(S_t = j \mid \tilde{\mathbf{r}}_t)}$$

With each iteration we obtain $\Pr(S_t = 1 \mid S_{t+1}, \tilde{\mathbf{r}}_t)$, and draw S_t.

Step 3: Now based on the state variables $\tilde{\boldsymbol{\beta}}_T = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_T]'$ from Step 1, the state $\tilde{S}_T = [S_1\ S_2\ ...\ S_T]'$ from Step 2, and the data $\tilde{\mathbf{r}}_T = [\mathbf{r}_1\ \mathbf{r}_2\ ...\ \mathbf{r}_T]'$, we generate the rest of the parameters $\{\boldsymbol{\Sigma}, \omega_0, \omega_1, \sigma_{u0}^2, \sigma_{u1}^2, \sigma_{v0}^2, \sigma_{v1}^2, \sigma_{uv0}, \sigma_{uv1}, \phi_1, \phi_2, p, q\}$ in this final step.

3.1: Based on the state $\tilde{S}_T = [S_1\ S_2\ ...\ S_T]'$, we generate the transition probabilities by

$$p \mid \tilde{S}_T \sim beta(u_{11} + n_{11}, u_{10} + n_{10})$$

$$q \mid \tilde{S}_T \sim beta(u_{00} + n_{00}, u_{01} + n_{01})$$

where $u_{11}/(u_{11} + u_{10})$ is our prior for p and $u_{00}/(u_{00} + u_{01})$ is our prior for q, and n_{ij} in the number of transitions from state i to j calculated from $\tilde{S}_T = [S_1\ S_2\ ...\ S_T]'$.

Next we generate the premia $\{\omega_0, \omega_1\}$. Rewriting the measurement equation as: $\mathbf{r}_t = \omega_0 + (\omega_1 - \omega_0)S_t + \mathbf{T}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$, and for each yield divide both sides with the standard deviation of the measurement error to obtain

$$\boldsymbol{\Sigma}^{1/2}\mathbf{r}_t - \boldsymbol{\Sigma}^{1/2}\mathbf{T}\boldsymbol{\beta}_t = \boldsymbol{\Sigma}^{1/2}\omega_0 + \boldsymbol{\Sigma}^{1/2}(\omega_1 - \omega_0)S_t + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}_t = \mathbf{X}\omega + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}_t \text{ with } \omega = \begin{bmatrix} \omega_0 \\ \omega_1 - \omega_0 \end{bmatrix}.$$

Given the prior for the premia $\omega \mid \boldsymbol{\Sigma} \sim N(\mathbf{b}_0, \mathbf{B}_0)$, we obtain the posterior distribution:

$\omega \mid \mathbf{\Sigma}, \tilde{S}_T, \tilde{\boldsymbol{\beta}}_T, ... \sim N(\mathbf{b}_1, \mathbf{B}_1)$ with $\mathbf{b}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'(\mathbf{\Sigma}^{1/2}\mathbf{r}_t - \mathbf{\Sigma}^{1/2}\mathbf{T}\boldsymbol{\beta}_t))$ and $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}$.

3.2: Now we generate the two AR(2) coefficients by rewriting the cycle as: $\frac{c_t}{\sigma_{vS_t}} = \frac{\phi_1 c_{t-1}}{\sigma_{vS_t}} + \frac{\phi_2 c_{t-2}}{\sigma_{vS_t}} + \frac{v_t}{\sigma_{vS_t}}$, or to put it in matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\varphi} + \mathbf{v}$, where $\boldsymbol{\varphi} = [\phi_1 \quad \phi_2]'$. Given a prior distribution for $\boldsymbol{\varphi} = [\phi_1 \quad \phi_2]'$ as $\boldsymbol{\varphi} \mid \sigma^2_{vS_t} \sim N(\mathbf{b}_0, \mathbf{B}_0)$, the posterior distribution is obtained as: $\boldsymbol{\varphi} \mid \sigma^2_{vS_t}, \tilde{S}_T, \tilde{c}_T \sim N(\mathbf{b}_1, \mathbf{B}_1)$ where $b_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{Y})$ and $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}$. We reject draws with non-stationary and unstable roots, and also when the two coefficients sum up to be more than 0.95.

3.3: For the errors in the measurement equation, we begin with the prior distribution for the variance for each of the yield i: $\sigma^2_{\varepsilon i} \sim IG(v_i/2, f_i/2)$, which is the inverse-Gamma distribution. The posterior distribution is defined as follows:

$$\sigma^2_{\varepsilon i} \mid \phi_1, \phi_2, \tilde{\mathbf{r}}_T, \tilde{\boldsymbol{\beta}}_T, \boldsymbol{\theta}_{S_t} \sim IG\left((v_i + T - 2)/2, \left(f_i + (\tilde{\mathbf{r}}_T - \boldsymbol{\theta}_{S_t} - \mathbf{T}\boldsymbol{\beta}_t)'(\tilde{\mathbf{r}}_T - \boldsymbol{\theta}_{S_t} - \mathbf{T}\boldsymbol{\beta}_t)\right)/2\right).$$

Finally we generate the errors for the state equations. Rewriting the first two of the state variables in matrix form as $\boldsymbol{\beta}^*_t = \mathbf{H}^*\boldsymbol{\beta}_{t-1}^* + \mathbf{n}^*_t$, $\mathbf{n}_t^* \sim N\left(\mathbf{0}, \mathbf{Q}_0^{*1/2}(\mathbf{I} + S_t\mathbf{\Gamma})\mathbf{Q}_0^{*1/2}{}'\right)$,

$$(\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \boldsymbol{\beta}_t^* = (\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \mathbf{H}^*\boldsymbol{\beta}_{t-1}^* + (\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \mathbf{n}_t^*$$

Given the prior distribution $\mathbf{Q}_0^* \mid \phi_1, \phi_2, \mathbf{\Gamma} \sim W(v_0/2, f_0/2)$, which is the Wishart distribution to allow for a covariance term. The posterior distribution is

$$\mathbf{Q}_0^* \mid \phi_1, \phi_2, \mathbf{\Gamma}, \tilde{S}_T, \tilde{\boldsymbol{\beta}}_t^* \sim W\left((v_0 + T)/2, \begin{pmatrix} f_0 + \left((\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \boldsymbol{\beta}_t^* - (\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right)' \\ \left((\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \tilde{\boldsymbol{\beta}}_t^* - (\mathbf{I} + S_t\mathbf{\Gamma})^{1/2} \mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right) \end{pmatrix} /2\right)$$

Next we have

$\mathbf{Q}_0^{*1/2}\boldsymbol{\beta}_t^* = \mathbf{Q}_0^{*1/2}\mathbf{H}^*\boldsymbol{\beta}_{t-1}^* + \mathbf{Q}_0^{*1/2}\mathbf{n}_t^*$, given a prior distribution

$\mathbf{I} + \boldsymbol{\Gamma} \mid \mathbf{Q}_0^*, \phi_1, \phi_2 \sim W\left(v_1/2, f_1/2\right)$, we have the posterior:

$$\mathbf{I} + \boldsymbol{\Gamma} \mid \mathbf{Q}_0^*, \phi_1, \phi_2, \tilde{S}_T, \tilde{\boldsymbol{\beta}}_t^* \sim W\left(\left(v_1 + T_1\right)/2, \begin{pmatrix} f_1 + \left(\mathbf{Q}_0^{*1/2}\boldsymbol{\beta}_t^* - \mathbf{Q}_0^{*1/2}\mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right)' \\ \left(\mathbf{Q}_0^{*1/2}\tilde{\boldsymbol{\beta}}_t^* - \mathbf{Q}_0^{*1/2}\mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right) \end{pmatrix} / 2\right)$$

Notice that T_1 is the number of times $S_t = 1$ and the product $\left(\mathbf{Q}_0^{*1/2}\boldsymbol{\beta}_t^* - \mathbf{Q}_0^{*1/2}\mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right)'\left(\mathbf{Q}_0^{*1/2}\tilde{\boldsymbol{\beta}}_t^* - \mathbf{Q}_0^{*1/2}\mathbf{H}^*\boldsymbol{\beta}_{t-1}^*\right)$ is only counted for observations at which $S_t = 1$.

Steps 1 to 3 complete one iteration of the Gibbs sampling procedure. We perform 10000 iterations, and keep the last 8000 for inference. In the Gibbs sampling procedure for the baseline model, we only omit Step 2 and simplify other relevant sections (e.g. only one set of term premia is drawn).