

Predictive Macro-Finance With Dynamic Partition Models

Daniel ZANTEDESCHI, Paul DAMIEN, and Nicholas G. POLSON

Dynamic partition models are used to predict movements in the term structure of interest rates. This allows one to understand historic cycles in the performance of how interest rates behave, and to offer policy makers guidance regarding future expectations on their evolution. Our approach allows for a random number of possible change points in the term structure of interest rates. We use particle learning to learn about the unobserved state variables in a new class of dynamic product partition models that relate macro-variables to term structures. The empirical results, using data from 1970 to 2000, clearly identifies some of the key shocks to the economy, such as recessions. We construct a time series of Bayes factors that, surprisingly, could serve as a leading indicator of economic activity, validated via a Granger causality test. Finally, the in-sample and out-of-sample forecasts from our model are quite robust regardless of the time to maturity of interest rates.

KEY WORDS: Bayes factor; Bond yield; Nelson–Siegel model; Particle learning.

1. INTRODUCTION

The recent economic crisis has brought into focus the importance of having sound methods to understand the behavior of financial and macro-economic time series. We use the term “predictive macro-finance” to describe statistical models that use macro-economic variables to predict financial asset returns. Understanding the term structure, or yield curves, of different bond maturities has a long history. Recently the Nelson–Siegel (1987) yield curve model (N–S) has been rekindled by a series of articles showing superior predictive performance over traditional models. The articles of Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006) highlight the three factor N–S model as a powerful tool for forecasting changes in rates. These authors showed that their model-based forecasts outperform many of the standard models. Koopman, Mallee, and van der Wel (2010) used a time-varying loadings version to achieve further gains.

We extend the dynamic N–S approach in a number of ways. First, we propose a dynamic product partition model (PPM) to allow for an unknown number of change points. This addresses one of the fundamental challenges in yield curve modeling where parameter shifts occur for various reasons, either policy-induced (e.g., the Volcker period of the early 1980s) or purely probabilistic (e.g., the 1995 change point reported in Chib and Kang 2009). Our approach incorporates both of these effects. Second, we develop a particle learning strategy to learn about the unobserved state variables and parameters. Third, our empirical results identify many of the key shocks (e.g., recessions, expansions) to the economy. We quantify the gains from our extensions from both in-sample fit and out-of-sample forecast perspectives. We also show that our results are robust to the maturity of interest rates. Finally, and most importantly, we construct a time series of regime diagnostics based on Bayes factors that can serve as a leading indicator of economic activity, validated via a Granger causality test.

Fama (2006) observed that if spot rates are subject to several permanent shocks to the system, then limiting the number of regime changes up front to two or three may prove inadequate to explain the data. Our approach addresses this by allowing for a possible regime change at every time point, and our analysis assesses the posterior probability of possible change points given the dynamic N–S term structure model. Two alternative approaches include those of Sims and Zha (1998, 2006), who modeled regime switches in U.S. monetary policy with a fixed number of (at most nine) regimes. Chib and Kang (2009) used up to four change points by exploiting a transient parameterization of a hidden Markov Model.

One challenging issue when fitting dynamic state-space models with a random number of change points is the computational explosion of the number of possible models. We argue that PPMs (Barry and Hartigan 1992, 1993), when coupled with particle learning methods, bypass this dilemma. Our approach also calculates a sequential Bayes factor at each time point for model comparison. We thus allow for multiple, random change points—a first in the macro-finance context. Table 1 compares our approach with that of competing models in the literature.

Explaining the cross-section of rates has traditionally been achieved with affine term-structure models (ATSMs; see, e.g., Chib and Ergashev 2009). Theoretically, these models are very popular because of their arbitrage-free properties; however, they have suffered somewhat in out-of-sample time series prediction exercises (see, e.g., Duffee 2002). The Bank of International Settlements (2005) reported that 9 out of 13 central banks that report their term structure modeling results to the BIS use the N–S model (or some minor variation thereof) rather than ATSMs to estimate yield curves.

The rest of the article is organized as follows. Section 2 provides a general framework for term structure models with macro-economic factors using the N–S model. Section 3 shows how to extend these models to allow for multiple change points in the form of a PPM. Section 4 details the term structure and macro-finance monthly data for 1970–2000 used in this study, which is used in Section 5 to carry out a comprehensive analysis. Section 6 concludes.

Daniel Zantedeschi is Graduate Researcher (E-mail: daniel.zantedeschi@mcombs.utexas.edu) and Paul Damien is Professor (E-mail: paul.damien@mcombs.utexas.edu), McCombs School of Business, University of Texas at Austin, 1 University Station B6000, Austin, TX 78712. Nicholas G. Polson is Professor, Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637 (E-mail: nicholas.polson@ChicagoBooth.edu).

Table 1. Main predictive macro-finance model presented in the literature

Authors	Model	Switching	N	TV	ESTIM	Link
Bansal and Zhou (2002)	ATSM	HMM	2	NO	GMM	
Dai and Singleton (2007)	ATSM	HMM	2	NO	MLE	
Ang and Bekaert (2002)	ATSM	HMM	2	NO	MLE	BC
Chib and Kang (2009)	ATSM	CH-HMM	4	NO	MCMC	
Zhu and Rahman (2009)	NS	HMM	2	NO	MCMC	
Sims and Zha (2006)	VAR	HMM	9	YES	MCMC	
Koopman, Mallee, and van der Wel (2010)	NS-G	NO		YES	MLE	
Diebold and Li (2006)	NS	NO		YES	MLE	
Diebold, Li, and Yue (2008)	NS	NO		YES	MLE-MCMC	BC
Christensen, Diebold, and Rudebusch (2007)	A-NS	NO		YES	MLE	
Ang and Piazzesi (2003)	ATSM	NO		YES	MCMC	MP
Mumtaz and Surico (2009)	NS	NO		YES	MCMC	MP
Our proposal	NS+	PPM	∞	YES	PL-MCMC	BC

NOTE: ATSM: Affine term structure model, NS: Nelson-Siegel model, A-NS: Affine version of Nelson-Siegel, NS-G: Nelson-Siegel with garch effects, NS+: Can be extended to ATSM, VAR: Vector auto-regression, N : Number of regimes or change points assumed, HMM: Hidden Markov model, CH-HMM: Transient hidden Markov model as in Chib and Kang (2009), PPM: Product partition model, TV: Time varying parameters, GMM: Generalized method of moments, MLE: Maximum likelihood estimation, MCMC: Markov chain Monte Carlo, PL-MCMC: Particle learning plus MCMC, Link: Type of bi-directional analysis, BC: Business cycle, MP: Monetary policy.

2. DYNAMIC MACRO NELSON-SIEGEL MODEL

Dynamic arbitrage free models demonstrate that three factors related to level, slope, and curvature drive yields and one-period returns. The dynamic N-S yield curve model comprises three principal factors: level, slope, and curvature. In addition, a fourth time-varying parameter can be used to determine the speed of time decay for the slope and curvature parameters. Suppose that we observe data on a panel of bond yields, $Y_{t+1}(\tau)$, at time $t + 1$ for each maturity τ with maturities indexed from $m = 1, \dots, M$. Diebold, Rudebusch, and Aruoba (2006) and Ang and Piazzesi (2003) introduced macro-economic factors in the evolution equation and provided theoretical as well as predictive support for the model. Our dynamic state-space specification of the N-S model with macro factors takes the form:

$$\begin{cases} Y_{t+1}(\tau) = \mathbf{F}(\lambda_{t+1})\mathbf{B}_t + \epsilon_{t+1}, \\ \theta_{t+1} = \mathbf{G}\theta_t + \mathbf{D}\kappa_t + \omega_{t+1}, \end{cases}$$

$$\begin{cases} \epsilon_{t+1} \\ \omega_{t+1} \end{cases} \sim \mathcal{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{E} & \\ & \mathbf{W} \end{pmatrix} \right]. \quad (1)$$

The vector \mathbf{B}_t contains the three $\beta_{.,t}$ loadings. The three parameters $\beta_{.,t}$ are the level, slope, and curvature of the yield curve. We use λ_t to model the decay in the term structure representing the fourth factor. The matrix $\mathbf{F}(\lambda_{t+1})$ contains the usual N-S functional specifications, and its (i, j) component is given by

$$\mathbf{F}_{i,j}(\lambda_{t+1}) = \begin{cases} 1, & j = 1 \\ \left(1 - \exp\left(-\frac{\tau_i}{\exp(\lambda_{t+1})}\right)\right) / \frac{\tau_i}{\exp(\lambda_{t+1})}, & j = 2 \\ \left(1 - \exp\left(-\frac{\tau_i}{\exp(\lambda_{t+1})}\right)\right) / \frac{\tau_i}{\exp(\lambda_{t+1})} - \exp\left(-\frac{\tau_i}{\exp(\lambda_{t+1})}\right), & j = 3. \end{cases}$$

The vector $\kappa_t = [1, k_{1,t}, k_{2,t}, k_{3,t}, k_{4,t}]'$ contains the intercept and four macro factors at time t . Our parameterization of the

decay parameter is different than that presented by Koopman, Mallee, and van der Wel (2010).¹ Consequently, we have two classes of parameters:

- $\theta_t = [\mathbf{B}_t, \lambda_t] = [\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \lambda_t]'$ vector of states determining the evolution of the N-S term structure that are time-varying.
- $\eta = [\mathbf{E}, \mathbf{W}, \mathbf{G}, \mathbf{D}]$. Here \mathbf{G} is a 4×4 matrix of state dynamics, \mathbf{D} is a 4×5 matrix of the (lagged) contribution of intercept and macro factors κ_t to the states θ_{t+1} . \mathbf{W} is the variance-covariance matrix for the state error term ω_{t+1} and, similarly, \mathbf{E} is the variance-covariance matrix for the observation error term ϵ_{t+1} . For notational simplicity, because η parameters are time-dependent but not time-varying, we drop the subscript t .

The form of the yield curve is determined by the three components and their associated weights, $\beta_{.,t}$. The first component takes the value 1 (constant) and thus can be interpreted as the overall level that influences the short and long term interest rates equally. The second component converges to 1 as $\tau \rightarrow 0$ and to 0 as $\tau \rightarrow \infty$ for a given t ; this influences mainly short-term interest rates. The third component, associated with medium-term rates, converges to 0 as $\tau \rightarrow 0$ and as $\tau \rightarrow \infty$ but is concave in τ , for a given t . Because the first component is the only component that equals 1 as $\tau \rightarrow \infty$, its corresponding $\beta_{1,t}$ coefficient is usually linked with the long-term interest rate. By defining the slope $y_t(\tau_\infty) - y_t(\tau_0)$, that it converges to $-\beta_{2,t}$ for a given t can be easily verified. Finally, the shape of the yield curve can be defined as $|y_t(\tau_i) - y_t(\tau_0)| - |y_t(\tau_\infty) - y_t(\tau_i)|$ for a medium maturity τ_i , whereas this shape converges to $\beta_{3,t}$ for a given t .

¹ They parameterized $Y_t(\tau_i) = \beta_{1,t} + \beta_{2,t}(\frac{1-\exp(-\tau_i\phi)}{\tau_i\phi}) + \beta_{3,t}(\frac{1-\exp(-\tau_i\phi)}{\tau_i\phi} - \exp(-\tau_i\phi)) + \epsilon_t$. Here ϕ is either fixed or is estimated through a stepwise deterministic spline in the range of 0-1 for numerical stability. We adopt a negative logarithmic transformation, $\lambda = -\log(\phi)$, to allow sampling of the decay parameter over the entire real line.

3. DYNAMIC PARTITION MODELS

PPMs were introduced by Barry and Hartigan (1992, 1993) as a clustering device producing time-inhomogenous time series. Barry and Hartigan articulated the following key idea:

These [PPM] models apply with special computational simplicity to change point problems, where the partitions divide the sequence of observations into components within which different regimes hold... [and] that the observations can eventually determine approximately the true partition.

Let $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ be a time series that we wish to partition into B contiguous segments defined by $\{1, \dots, t_1\}$, $\{t_1 + 1, \dots, t_2\}$, $\{t_{B-1}, \dots, T\}$ with t_1, t_2, \dots, t_{B-1} denoting the change points. Equivalently, we can use a T -dimensional set of binary indicators, $U = (1, u_2, \dots, u_{T-1}, 1)'$, with the event $u_t = 1$ implying a change point and $u_t = 0$ implying no change point.

PPMs use a random number of segments B obtained via a partition $\rho \equiv (1, t_1, \dots, t_{B-1}, T)$, which divides the data \mathbf{Y} into B contiguous blocks, denoted by $\mathbf{Y}_{t_{r-1}, t_r} = (\mathbf{y}_{t_{r-1}+1}, \dots, \mathbf{y}_{t_r})'$ for $r = 1, \dots, B-1$. Following Yao (1984) and Loschi and Cruz (2005), we measure the degree of similarity among the block of observations $\mathbf{Y}_{i,j}$ via a prior cohesion distribution $c_{i,j}(p)$. This distribution influences the partition structure by setting the transition probabilities at the change points. Here p is the probability that a change point occurs at any instance in the sequence. The prior cohesion for the block $\mathbf{Y}_{i,j}$ is then given for all i, j , $i < j$ by

$$c_{i,j}(p) = \begin{cases} p(1-p)^{j-i-1} & \text{if } j < T \\ (1-p)^{j-i-1} & \text{if } j = T, \end{cases}$$

that is, the probability that a new change takes place after $j-i$ instants, given that a change occurred at the instant i . This implies that a sequence of change points establishes a discrete renewal process, where the occurrence times are identically distributed with a geometric rate depending on the length of the current partition.

With the foregoing development, we define $\{\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T); \rho\} \sim PPM$ if for every partition, the distribution of ρ is given by $P(\rho|p) = p^{b-1}(1-p)^{T-b}$. The conditional prior distribution for B is then $P(B=b|p) = \binom{T-1}{b-1} p^{b-1}(1-p)^{T-b}$ for $1 \leq b \leq B$. Conditional on ρ and p , the time series Y is independent of p . We now show how to incorporate a PPM structure into the nonlinear N-S yield curve model.

Given a partition structure ρ and cohesion function c , the likelihood function of observing a time series \mathbf{Y} of yields from the observation equation factors into a joint distribution, $P(\mathbf{Y}_\rho, \theta|\mathbf{E}, c)$, over all of the random variables given by

$$P(\mathbf{Y}_\rho, \theta|\mathbf{E}, c) = \left\{ \prod_{t_j \in \rho} \left(\prod_{i=t_{j-1}}^{t_j} P(\mathbf{Y}_{t_{j-1}, t_j}|\theta_i) p(\theta_i) \right) \times c_{t_{j-1}, t_j}(p) \right\} \times \pi(p),$$

where $p(\theta_i) = \int p(\theta_i|\mathbf{E}) p(\mathbf{E}) d\mathbf{E}$. Here $\pi(p)$ defines the prior probability of a change point.

The data factor, $\mathcal{P}_{[i,j]}(\mathbf{Y}_{[i,j]})$, for each $[i,j]$ block is given by

$$\mathcal{P}_{[i,j]}(\mathbf{Y}_{[i,j]}) = \int \int \left(\prod_{i=t_{j-1}}^{t_j} p(\mathbf{Y}_{t_{j-1}, t_j}|\theta_i) p(\theta_i) \right) \times c_{t_{j-1}, t_j}(p) \pi(p) dp d\theta,$$

where we have integrated over all of the uncertainties. We also use the following sequential factor, defined by $\mathcal{P}_{[t_{j-1}, t_j]}(\mathbf{Y}_{[t_{j-1}, t_j]}|u_{t_j} = 1)$ when $u_{t_j} = 1$ indicates a change point at time t_j , and $\mathcal{P}_{[t_{j-1}, t_j]}(\mathbf{Y}_{[t_{j-1}, t_j]}|u_{t_j} = 0)$ with no change point, given a set of restart parameters $\theta_{t_j}^*$. Hereinafter, without loss of generality, we denote $t_j \equiv t$, $t_{j+1} \equiv t+1$, and so on, for any j .

Sequential Change Point Probabilities

The set of binary indicators U identifies the total number of model parameters via $B_U = 1 + \sum_{t=1}^{T-1} (1 - u_t)$. Under the prior $\pi(p) \sim \text{Beta}(\alpha, \beta)$, we have

$$\frac{\int_0^1 p^{b_t-2} (1-p)^{t-b_t+1} \pi(p) dp}{\int_0^1 p^{b_t-1} (1-p)^{t-b_t} \pi(p) dp} = \frac{\Gamma(t+\beta-b_t-1)\Gamma(b_t+\alpha)}{\Gamma(b_t+\alpha-1)\Gamma(t+\beta-b_t)} = \frac{(\alpha+b_t-1)}{(t-b_t-1+\beta)},$$

where b_t denotes the number of change points between times 1 and t .

The posterior distribution of p , at time t , is Bernoulli with probability proportional to

$$\Pi_t = \frac{\mathcal{P}_{[t,t+1]}(Y_{[t,t+1]}|u_t = 1)}{\mathcal{P}_{[t,t+1]}(Y_{[t,t+1]}|u_t = 0)} \frac{(\alpha+b_t-1)}{(t-b_t-1+\beta)}.$$

This is simply the probability ratio of a change point to no-change point based on the PPM structure. We complement these data factors with the corresponding predictive densities.

4. SEQUENTIAL MONTE CARLO FOR TERM STRUCTURE MODELING

Sequential Monte Carlo (SMC) is an alternative to traditional Markov chain Monte Carlo (MCMC) methods that is designed for online inference in dynamic, possibly nonlinear models. It allows us to provide the full set of filtering distributions together with the sequential posterior distributions of the parameters of interest. Specifically, at time t , we have a set of random draws, or particles $\{Z_t^{(\gamma)}\}_{\gamma=1}^N$, which has a filtering distribution $p(Z_t|Y_t)$ that contains the sufficient information about all of the uncertainties given the data up to time t . Given the next observation, Y_{t+1} , the key task is to update the particles from t to $t+1$ and provide draws from the next filtering distribution. Carvalho et al. (2010) recommended particle learning (PL) for filtering nonlinear state-space models. The next filtering distribution is

$$p(Z_{t+1}|Y^{t+1}) = \int p(Z_{t+1}|Z_t, Y_{t+1}) d\mathbb{P}(Z_t|Y^{t+1}) \propto \int p(Z_{t+1}|Z_t, Y_{t+1}) p(Y_{t+1}|Z_t) d\mathbb{P}(Z_t|Y^t),$$

where $p(Y_{t+1}|Z_t)$ and $p(Z_{t+1}|Z_t, Y_{t+1})$ denote the predictive and propagation rules, respectively, and $Y^t = (Y_1, \dots, Y_t)$. This suggests the following particle approximation:

- Resample indices with replacement from a multinomial distribution where each index has weight $w_\gamma \propto p(Y_{t+1}|Z_t^{(\gamma)})$.
- Propagate with a draw from $Z_{t+1} \sim p(Z_{t+1}|Z_t^\gamma, Y_{t+1})$ to obtain a new collection of particles $\{Z_{t+1}^{(\gamma)}\}_{\gamma=1}^N \sim p(Z_{t+1}|Y^{t+1})$.

We now show how to define and update particles for Z_t in the dynamic N-S model.

4.1 Sequential Product Partition Nelson–Siegel Model

Our algorithm consists broadly of four phases: (1) predictive resampling, (2) propagation, (3) updates of states and sufficient statistics, and (4) updates of the fixed parameters. These are preceded by an initialization phase, which differs between in-sample and out-of-sample calculations. A graphical illustration of the entire model is given in Figure 1. Appendix A presents details on priors and computation. Now consider the sequential update for time $t + 1$ given the new data Y_{t+1} .

The following quantities are available due to the updating at time t : the mean and covariance, f_t, Q_t , of the predictive density with no change point; f_t^*, Q_t^* , the mean and covariance of the predictive density under a change point; and $\{\theta_t^\gamma, s_t^{\theta, \gamma}, \eta^\gamma\}_{\gamma=1}^N$ containing the N–S factors, namely θ_t^γ , their sufficient statistics, $s_t^{\theta, \gamma} = (m_t, C_t)^{\theta, \gamma}$, and the static parameters η^γ .

Sufficient Information: With the foregoing definitions, we track $\{Z_t^\gamma\}_{\gamma=1}^N = \{\theta_t^\gamma, s_t^{\theta, \gamma}, \eta^\gamma\}_{\gamma=1}^N$.

Predictive Distribution: To perform predictive resampling, we use the predictive distribution of the measurement equation. Under the usual conditionally conjugate Normal-Wishart prior, the predictive density follows a multivariate *student-t* distribution: $\mathcal{P}_{[t,t+1]}(Y_{[t,t+1]}|u_t = 0) \sim \mathcal{MVT}(f_t, Q_t, M - 3)$, with f_t and Q_t defined above.

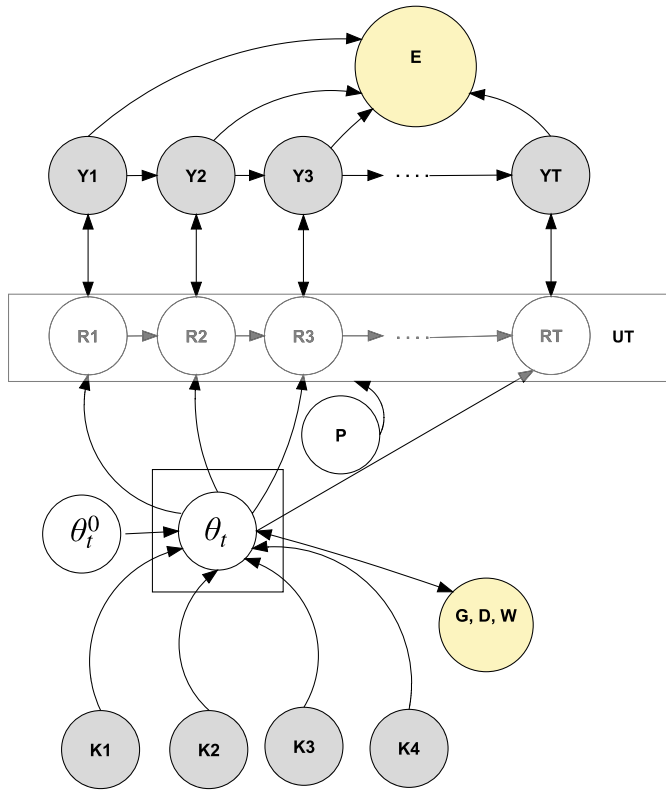


Figure 1. Model structure and parameters as a Bayesian network. The square plate indicates $T - 2$ sequential updates of θ_t . The K 's are macro factors. $\mathbf{G}, \mathbf{D}, \mathbf{W}$, and \mathbf{E} are fixed parameters, updated via Gibbs and Kalman filtering. The Y_i 's are the yields data. θ_t^0 is the initialization point. The R_i 's are possible regime changes implied by \mathbf{u} , and p is the prior probability of a change point. The online version of this figure is in color.

Predictive Resampling: For each particle, we calculate the data factor $\mathcal{P}_{[t,t+1]}$ together with resampling weights $\{w_{t+1}^\gamma\}_{\gamma=1}^N$. We then resample a set of particles $\{\theta_t^\gamma, s_t^{\theta, \gamma}, \eta^\gamma\}_{\gamma=1}^N$.

Propagation of Change Point Probabilities via Forward

Filtering: For each particle, we simulate a new u_t^γ with probability proportional to Π_t (defined earlier) in which $\mathcal{P}_{[t,t+1]}(Y_{[t,t+1]}|u_t = 0) \sim \mathcal{MVT}(f_t, Q_t, M - 3)$ with no change point at time t , and similarly with a change point at time t . m_t^* and C_t^* are given by the set of restarting points θ_t^* . Consider the sequence of latent variables $\{u_t^\gamma\}_{\gamma=1}^N$, and weights $\{w_{t+1}^\gamma\}_{\gamma=1}^N$. With no change point at time t , set $u_t^\gamma \equiv 0$, and thus propagate the new states $\{\theta_{t+1}^\gamma\}_{\gamma=1}^N$ from $\theta_{t+1} \sim \mathcal{MVN}(\mathbf{G}\theta_t + \mathbf{D}\kappa_t, \mathbf{W})$. With a change point at time t , set $u_t^\gamma \equiv 1$, and thus propagate the new states $\{\theta_{t+1}^\gamma\}_{\gamma=1}^N$ from $\theta_{t+1} \sim \mathcal{MVN}(\mathbf{G}\theta_t^* + \mathbf{D}\kappa_t, \mathbf{W})$.

Updating States and Sufficient Statistics: When there is no change point, $u_t^\gamma = 0$, we follow West and Harrison (1997) and define

$$\begin{aligned} a_{t+1} &= E(\theta_{t+1}|Y_{1:t}) = \mathbf{G}m_t + \mathbf{D}\kappa_{t+1}, \\ R_{t+1} &= \text{var}(\theta_{t+1}|Y_{1:t}) = \mathbf{G}C_t\mathbf{G} + \mathbf{W}, \\ f_{t+1} &= E(Y_{t+1}|Y_{1:t}) = \mathbf{F}'a_{t+1}, \\ Q_{t+1} &= \text{var}(Y_{t+1}|Y_{1:t}) = \mathbf{F}R_{t+1}\mathbf{F}' + \mathbf{E}. \end{aligned} \tag{2}$$

We obtain an updated mean of the states, $\hat{\theta}_{t+1} = \theta_{t+1} + R_{t+1}\mathbf{F}'C_{t+1}^{-1}(Y_{t+1} - \theta_{t+1})$.

We then update the sufficient statistics $s_{t+1}^\theta = (m_{t+1}, C_{t+1})$:

$$\begin{aligned} m_{t+1} &= E(\theta_{t+1}|Y_{t+1}) = a_{t+1} + R_{t+1}\mathbf{F}'Q_{t+1}^{-1}(Y_{t+1} - \theta_{t+1}), \\ C_{t+1} &= \text{var}(\theta_{t+1}|Y_{t+1}) = R_{t+1} - R_{t+1}\mathbf{F}'Q_{t+1}^{-1}\mathbf{F}R_{t+1}. \end{aligned}$$

With a change point at time t , $u_t^\gamma = 1$, and the update depends on the starting values at time t , (m_t^*, C_t^*) . The foregoing updating expressions are then modified simply by appropriately adding the superscript, $*$.

Updating the Fixed Parameters, η^γ : For each particle, compute $\hat{\mathbf{E}} = (Y - \mathbf{F}\theta_t^\gamma)'(Y - \mathbf{F}\theta_t^\gamma)$. The posterior distribution for $\mathbf{E} \sim \mathcal{IW}(t + M - 2, \hat{\mathbf{E}})$. Once θ_{t+1} has been updated, the state equation $\theta_{t+1} = \mathbf{G}\theta_t + \mathbf{D}\kappa_t + w_t$ can be reformulated as a vector autoregression. Again using a Jeffreys prior for \mathbf{W} and a Minnesota prior for \mathbf{G} and \mathbf{D} we can use a Gibbs sampler by sampling repeatedly from $\mathbf{W}|\theta, gd \sim \mathcal{IW}(v, \hat{\mathbf{S}})$ and $gd|\mathbf{W}, \theta \sim \mathcal{MVN}(gd, \hat{\Sigma}_{gd})$ with $gd = \text{vec}[\mathbf{G}, \mathbf{D}]$. Definitions of $v, \hat{\mathbf{S}}, \hat{gd}, \hat{\Sigma}_{gd}$ are provided in Appendix A.

Initialization: To initialize at time $t = 0$, we use ordinary least squares (OLS) for the in-sample data, following Diebold and Li (2006). In the out-of-sample exercise, we use the training period 1970–1974 and compute the initializing values consistently with the in-sample part.

5. TERM STRUCTURE DATA

The data comprise end-of-month yields for U.S. bonds between January 1970 and December 2000, a total of 372 observations. The monthly U.S. macroeconomic variables used in this

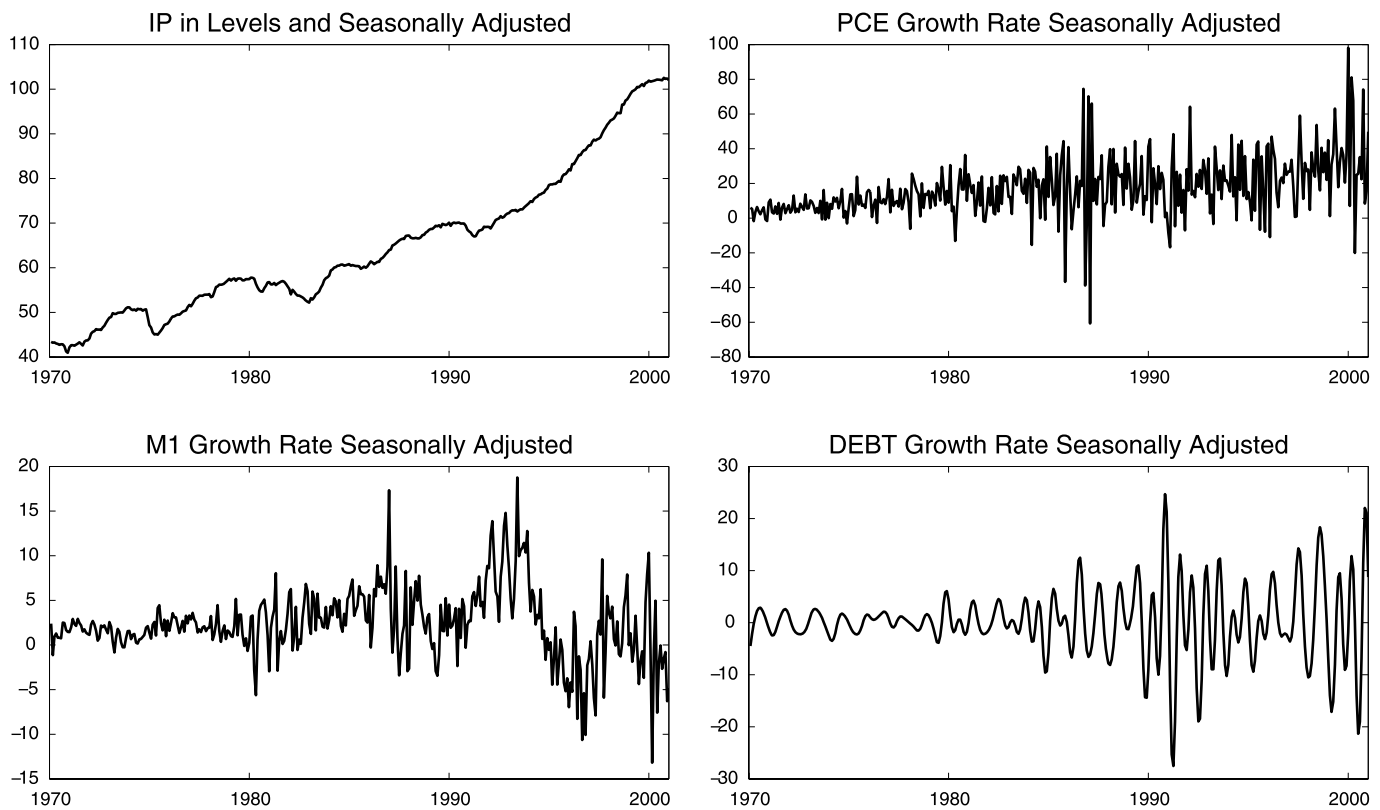


Figure 2. Macroeconomic variables, IP: industrial productivity, PCE: personal consumption expenditures, M1: monetary mass, DEBT: ratio between government spending and GDP.

work were obtained from the U.S. Federal Reserve Macroeconomic Database (FRED). These data consist of the unsmoothed Fama–Bliss yields (Fama and Bliss 1987; Bliss 1997) and were made available by Diebold and Li (2006). The data include linearly interpolated bond yields with maturities as in Diebold and Li (2006), along with the 1-month maturities. (In the interest of brevity, we refer the reader to the foregoing references for details on these data.) Bliss (1997) tested and compared five distinct methods for estimating the term structure. The unsmoothed Fama–Bliss method is an iterative method through which the discount rate function is built up by computing the forward rate necessary to price successively longer maturity bonds. The smoothed Fama–Bliss “smooths out” these discount rates by fitting an approximating function to the “unsmoothed” rates. Bliss argued that his tests establish the presence of unspecified, but nonetheless systematic, omitted factors in the prices of long maturity notes and bonds. Using parametric and nonparametric tests, he found that the unsmoothed Fama–Bliss did best overall (see also de Pooter, Ravazzolo, and van Dijk 2007). We include data from well before the Volcker period, consistent with Ang and Piazzesi (2003). We do this to (a) identify regime shifts over a long time horizon, (b) provide a large sample to better estimate the parameters of the model; and (c) assess the performance of the regime-shift model over periods with strikingly different characteristics. Panel (a) of Figure 3 (see Section 6.1) shows time series plots for the included yields.

5.1 Macroeconomic Data

Diebold, Rudebusch, and Aruoba (2006), Ang and Piazzesi (2003), and others have emphasized the importance of studying the impact of macro factors on the yield curve. For example, Diebold, Rudebusch, and Aruoba (2006) presented arguments supporting the introduction of these factors in the state or evolution equation of the dynamic version of the N–S model. Thus, consistent with the current literature detailed in Table 1, we considered monthly measures of inflation, economic activity, monetary policy, and fiscal policy. We used the inflation measure PCE (Personal Consumption Expenditures; chain-type price index). For economic activity, we relied on the series IP (Industrial Productivity). We included monetary policy through the the M1 monetary mass, and introduced fiscal policy (which is known to influence term structure) via the variable DEBT, which is quarterly fiscal policy interpolated to a monthly frequency.² Figure 2 depicts the four macro variable time series.

6. EMPIRICAL ANALYSIS

The macro variables are linked to the yields data via the dynamic N–S model. There are 18 yield curves, each corresponding to different maturity levels. A key empirical finding is that the regime shifts identified by our model coincide with actual events observed in the U.S. economy in the time frame 1970–2000. A second finding is that the regime shifts identified by

² Refer to <http://research.stlouisfed.org/fred2/> for details concerning variables units, levels, and seasonal adjustments.

the sequential PPM serve as a leading indicator of economic activity in the United States. Third, the in-sample and out-of-sample predictions are quite encouraging, regardless of time to maturity.

6.1 In-Sample Results

From Table 2, note that the average root mean squared error for the entire term structure is 0.0428, with a maximum of 0.0917 for the short-term rates. Comparing the top two panels in Figure 3 shows the excellent overall fit from the Bayesian analysis. The fit demonstrates the greatest fluctuation in the 1980s, with the first years of that decade enduring considerable interest rate volatility due to recession, inflation, and failure of monetary policy. The bottom two panels show the excellent fit for a subset of the data known as the Great Moderation from 1986 to 2000.

Figure 4 shows posterior estimates of (a) the coefficient θ_1 , (b) the slope θ_2 , and (c) the curvature θ_3 . Note that the level curve coincides nicely with the fluctuations in the entire term structure shown in Figure 3, and not just to the short-term rate.

For slope and curvature, Bliss (1997) suggested using the second and third principal components of the yields as “good” empirical estimators. Diebold and Li (2006) and Diebold, Li, and Yue (2008) provided a similar estimator, but fixed the decay parameter, whereas we leave it as stochastic. From Figure 4(b) and (c), note the positive correlation between curvature and slope shown by the similar paths. This is also demonstrated by the empirical estimates of the **G** and **D** matrices. The economics reasoning behind these positive correlations is detailed

Table 2. In-sample results obtained by Monte Carlo averaging along 10,000 particles of the algorithm

Maturity (months)	Yields mean	Pricing error mean in bp	Fitted yields st. dev.	RMSE
1	6.44	-2.3	2.319	0.0917
3	6.75	-1.5	2.407	0.0754
6	6.98	2.2	2.418	0.0587
9	7.10	0.3	2.407	0.0635
12	7.20	-1.4	2.393	0.0387
15	7.31	1.9	2.377	0.0537
18	7.38	0.1	2.364	0.0407
21	7.44	-2.5	2.351	0.0363
24	7.46	0.8	2.339	0.0162
30	7.55	1.1	2.316	0.0228
36	7.63	-2.5	2.296	0.0162
48	7.77	2.6	2.265	0.0264
60	7.84	-1.0	2.242	0.0281
72	7.96	-0.4	2.224	0.0322
84	7.99	1.1	2.212	0.0328
96	8.05	-1.5	2.202	0.0467
108	8.08	-1.4	2.196	0.0337
120	8.05	2.5	2.191	0.0567
Mean		-0.105	2.306	0.0428
Median		-0.007	2.322	0.0387

NOTE: Yields mean refers to the actual Fama–Bliss yields mean between 1970 and 2000. Pricing error mean in basis points refers to the difference between actual yields and average estimated yields; pricing error standard deviation in basis points and root mean squared error. A basis point, bp, is a unit relating to interest rates that is equal to 1/100th of a percentage point per annum.

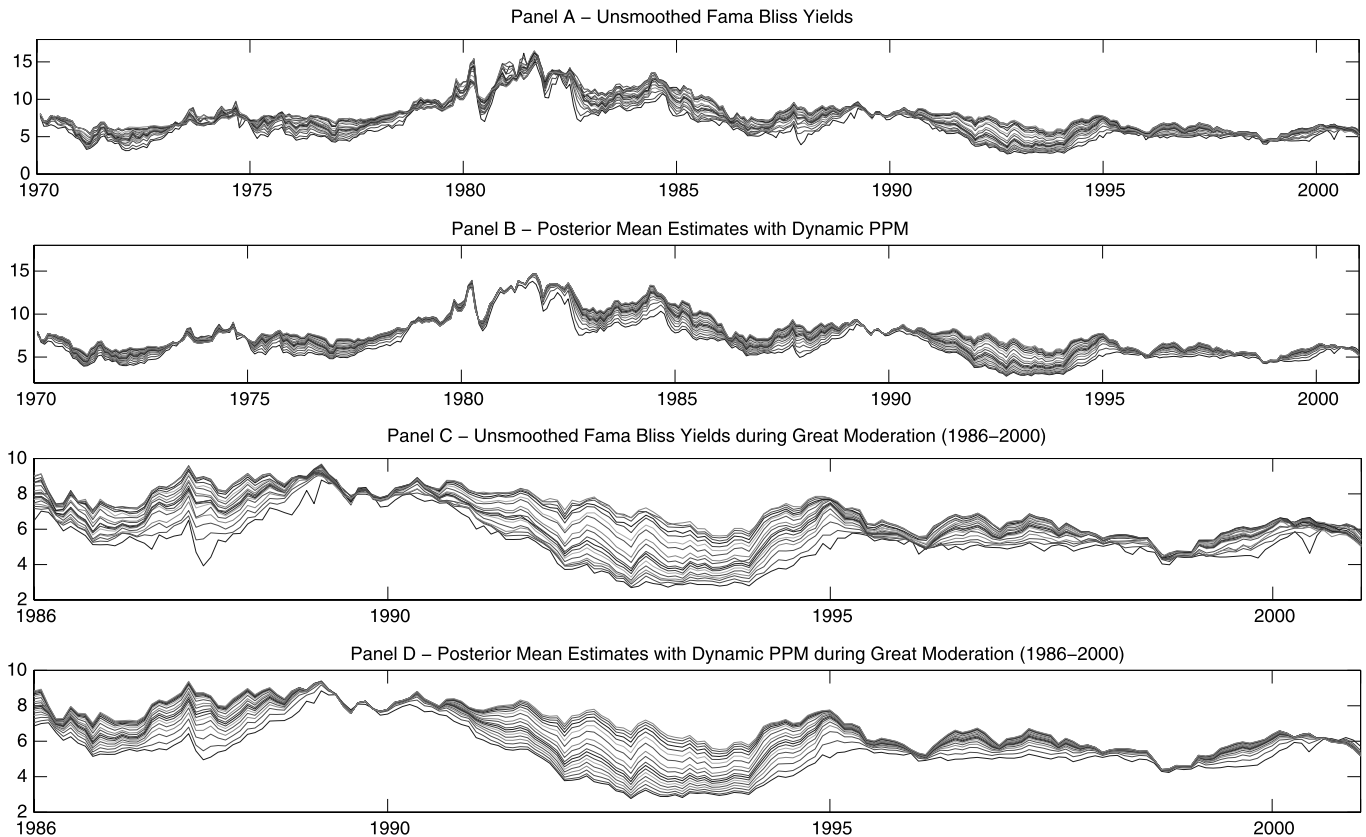


Figure 3. Comparison between actual Fama–Bliss yields and posterior mean averages during the entire time period, Panel A vs. Panel B; and during the Great Moderation period, Panel C vs. Panel D.

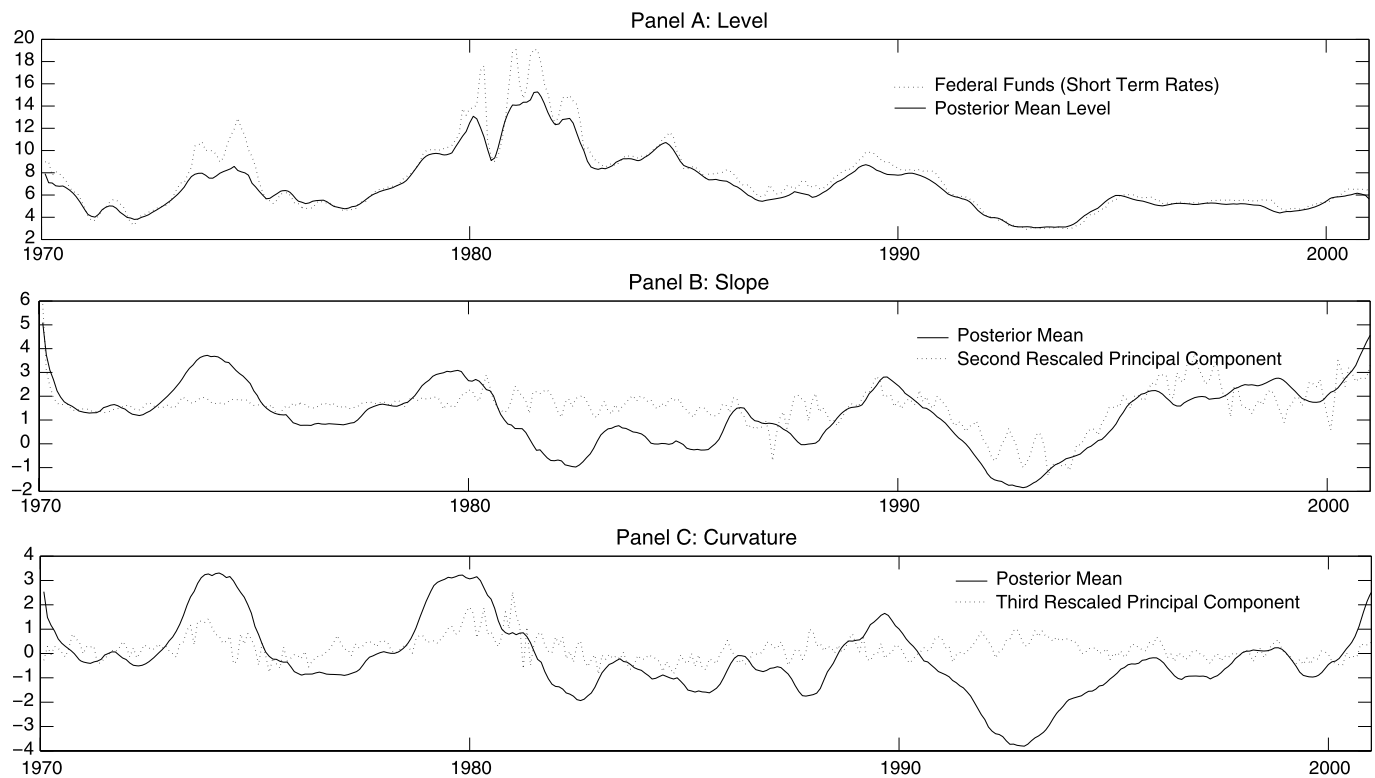


Figure 4. Panel A: Nelson–Siegel levels ($\theta_{1,t}$): Posterior mean, and the short term rate. Panel B: Nelson–Siegel slopes ($\theta_{2,t}$): Empirical estimate, and posterior mean. Panel C: Nelson–Siegel curvatures ($\theta_{3,t}$): Empirical estimate, and posterior mean.

in Section 6.4. The amount of output created from the Bayesian analysis is considerable, and we report only portions of the \mathbf{G} and \mathbf{D} estimates in Section 6.4.

6.2 Prior Assumptions on Change Point Probability p

To demonstrate the impact of the prior on the posterior probability of whether or not a time point is a regime change, Figure 5(a)–(c) depicts three different priors and the corresponding posterior distributions of p . Clearly, an informative prior would impact the distribution of change points. We use the uniform prior in our analysis; however, it is interesting to observe that the Bayesian approach allows an informed choice in the analysis where appropriate. Instead of providing a graph of probabilities of change and the most probable partitions, we later show the corresponding Bayes factor graph in Figure 6 that depicts the most probable regime changes.

Consider Table 3, in which we arbitrarily chose three event months in which we detected regime changes and one event month where there was no regime change. The table reports the posterior probabilities of the change points as well as the posterior probabilities of change points in the months preceding and succeeding the event month. As one example, consider the First Gulf War event. The posterior probability of regime change during August 1990 is 65.68%, and the posterior probabilities of change points in July and September are 36.13% and 51.36%, respectively. As we show later, shocks such as a recession or a war are not one-off events, and the effects of such events are likely to be carried over in the future (see Section 6.5).

6.3 Out-of-Sample Results

A common approach is to use rolling forecasts to evaluate out-of-sample performance. We train the model between 1970 and 1974 and then, starting in January 1975, calculate one-step-ahead, three-step-ahead, and six-step-ahead forecasts up to December 2000. We present our aggregate forecasts in Table 4, along with those from a standard random-walk model for comparison.

The macro finance literature shows that a random-walk model outperforms in the short run. Table 4 confirms this for our dataset showing one-step-ahead predictions for the 3-month and 1-year maturity levels. For all the six-step-ahead forecasts and longer maturities, our model dominates the random-walk model due in part to mean reversion. Our model compares favorably with other models in terms of prediction errors. In particular, at three-step-ahead and six-step-ahead forecasts, our predictions are much improved.

In Table 5, to assess model performance, we calculate the PPC of Gelfand and Ghosh (1998). For any model \mathcal{M}_j , the posterior PPC is defined as $PPC_j = D_j + W_j$, where

$$D_j = \frac{1}{M+2} \sum_{m=1, t=1}^{M, T} \text{var}(\tilde{y}_{m,t} | \mathcal{M}_j),$$

$$W_j = \frac{1}{M+2} \sum_{m=1, t=1}^{M, T} [y_{m,t} - E(\tilde{y}_{m,t} | \mathcal{M}_j)]^2,$$

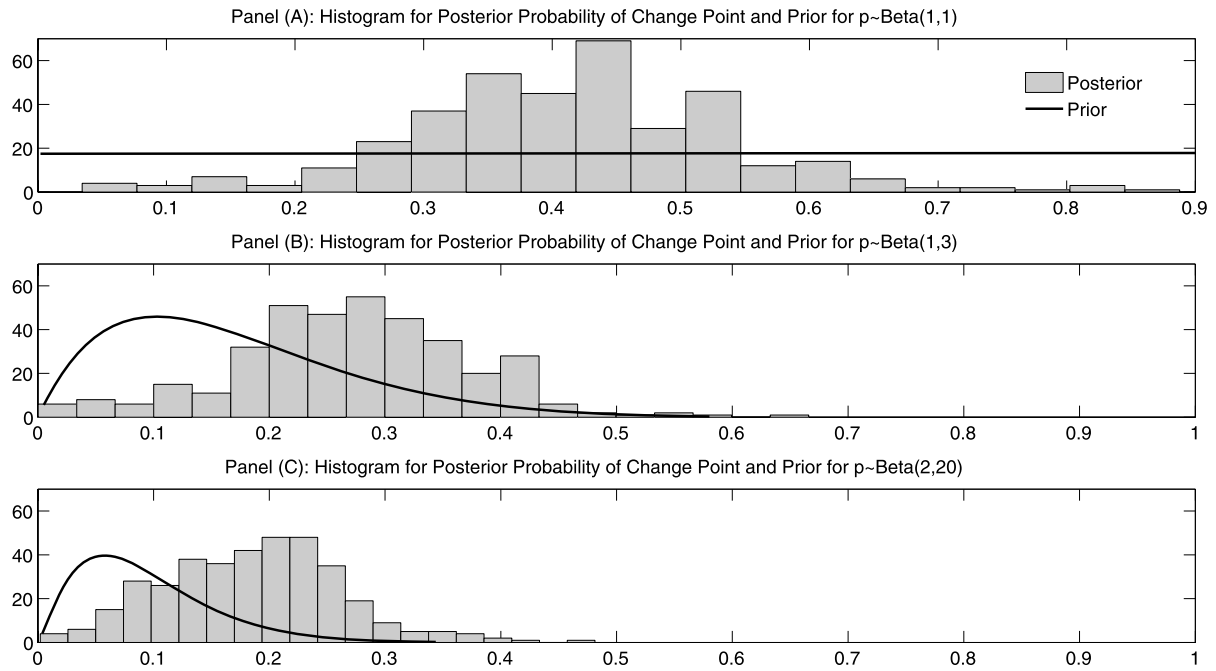


Figure 5. Prior to posterior sensitivity analysis for the changepoint probability. Summary statistics for the prior to posterior analysis: Panel A: prior $\mu = 0.5$, prior $\sigma = 0.2887$, posterior $\mu = 0.4142$, posterior $\sigma = 0.1319$; Panel B: prior $\mu = 0.25$, prior $\sigma = 0.1936$, posterior $\mu = 0.2706$, posterior $\sigma = 0.1015$; Panel C: prior $\mu = 0.0909$, prior $\sigma = 0.0599$, posterior $\mu = 0.1856$, posterior $\sigma = 0.0763$.

where $\tilde{y}_{m,t}$ denote the predictions of the yields for maturity m and time t under \mathcal{M}_j . The first term, D_j , is a penalty function that favors more parsimonious models, whereas the second term, W_j , measures goodness of fit. Smaller values of PPC are preferable. Based on this metric, our model outperforms the random-walk model for the three-step-ahead and six-step-ahead predictions, and is comparable for one-step-ahead forecasts. Table 5 also includes coverage probabilities based on 95% credible intervals, which demonstrate that our model is robust to this metric through all of the prediction steps.

6.4 Posterior Estimates for the Nelson–Siegel Model

Table 6 provides posterior mean estimates of the parameters in the state dynamics. We interpret “significant” to mean that “standard errors are sufficiently smaller in magnitude than the coefficients.” When the state variable is level or slope, the corresponding lagged coefficients are significant. The lags contain information about interest rates by virtue of the construction of a dynamic state-space model and the recursive learning between the observation and state equations. The lagged decay variable is significant in the slope and curvature regressions, but not in

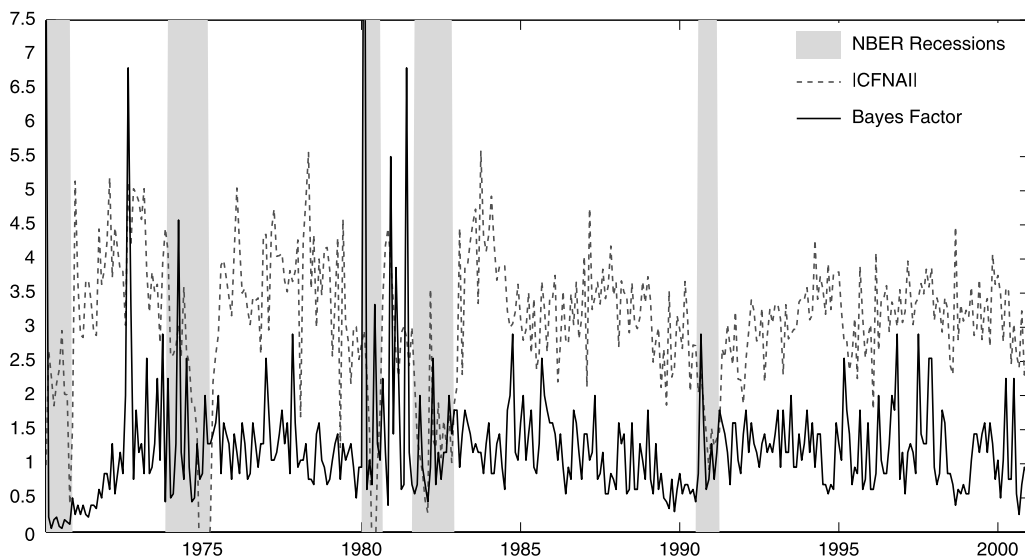


Figure 6. Bayes factors (solid curve); absolute value of CFNAI indicator (dashed curve), and recession regions (shaded areas).

Table 3. Sensitivity analysis to different priors for p with respect to four particular events

Event	Date	Post. prob.		
		Unif	Beta(1, 3)	Beta(2, 20)
Oil crisis	02/74	0.4607	0.3453	0.2336
	03/74	0.7511	0.4421	0.3660
	04/74	0.4350	0.2791	0.1724
Monetary	12/79	0.3854	0.2739	0.1205
	01/80	0.8879	0.6660	0.4814
	02/80	0.2920	0.1048	0.1407
Gulf war	07/90	0.3613	0.2587	0.1488
	08/90	0.6568	0.4174	0.3620
	09/90	0.5136	0.4069	0.1956
End-year	11/94	0.2920	0.1813	0.1282
	12/94	0.5136	0.4118	0.2363
	01/95	0.4607	0.3448	0.2284

NOTE: First oil crisis in 1974, monetary shock in 1980, First Gulf War in 1990, and a month at the end of 1994 that did not experience any particular shock apart from seasonal turbulence related to the end of the fiscal year. For each event, the value of the posterior probability for a time window of 3 months centered around the event is shown. Unif: Uniform prior, the one used throughout the analysis.

the level equation. This is compatible with the analysis of Duffee (2002, 2006).

Our macroeconomic variables appear to not influence the state dynamics directly, except for DEBT and the monetary

Table 4. Out-of-sample comparisons with the random-walk (RW) model

Maturity	Model				Our model
	RW	AR(1)	VAR(1)	NS	
Panel A: 1 month ahead					
3 months	0.18	-5	-1	+3	+1
1 year	0.23	-4	-2	+2	+1
3 years	0.25	-2	-2	+5	-3
5 years	0.29	-1	-1	+6	-2
10 years	0.28	-3	-1	+3	-1
Panel B: 3 months ahead					
3 months	0.39	-5	-6	+1	-2
1 year	0.52	-2	-4	+3	-6
3 years	0.53	-1	0	+1	-10
5 years	0.55	-1	-1	-1	-9
10 years	0.49	0	0	0	-9
Panel C: 6 months ahead					
3 months	0.74	-5	-3	-11	-18
1 year	0.78	+1	+2	-8	-16
3 years	0.95	0	+4	-4	-15
5 years	0.93	0	+3	-3	-17
10 years	0.90	-1	+1	-3	-17

NOTE: The numbers in the RW column are the root mean squared errors (RMSEs) in basis points. The numbers in the rest of the body of the table are the percentage deviations from these RMSEs. The plus (minus) number indicate that the particular model is worse (better) than the RW model in percentage terms. For example, consider the 10-year maturity/6-months ahead scenario under Panel C. The performances of AR(1), VAR(1), and NS, are roughly the same. In contrast, our model does somewhat better. All models in the comparisons include macro factors used in this article. The models compared are AR(1), VAR(1), and NS, corresponding to autoregressive, vector autoregressive, and Nelson-Siegel models, consistent with similar comparisons of Diebold and Li (2006). The bold numbers also isolate the distinctions with respect to the RW, the other models, and our model.

Table 5. Out-of-sample results by coverage probabilities and posterior predictive criterion (PPC)

Maturity	1 step ahead		3 steps ahead		6 steps ahead	
	NS	RW	NS	RW	NS	RW
3 months	0.87	0.95	0.72	0.69	0.88	0.44
1 year	0.89	0.93	0.86	0.66	0.86	0.45
3 years	0.99	0.96	0.89	0.60	0.70	0.57
5 years	0.95	0.98	0.91	0.82	0.80	0.53
10 years	0.96	0.99	0.92	0.86	0.87	0.65
PPC	1.99	1.84	2.76	3.51	3.11	6.35

NOTE: Out-of-sample forecasts on a rolling basis between January 1976 and December 2000. N-S, Nelson-Siegel; RW, random walk. We use 10,000 particles for the N-S model and 10,000 realizations of the RW model. The first five rows are the coverage probabilities of the 0.95 credible intervals. PPC is the PPC of Gelfand and Ghosh averaged over all maturities.

supply variable M1 in the level regression. The DEBT coefficient is negative, as expected because it is defined as the ratio between government spending and GDP. Therefore, when GDP increases, DEBT decreases, which occurs during an expansion when interest rates are usually higher. The opposite is true in a recession. Also, because DEBT captures the impact of GDP, it offsets industrial productivity, IP. As expected, M1 is positive, because increases in monetary mass (M1) are usually associated with steeper interest rates, in anticipation of higher inflation. Our analysis confirms that the relationship between slope and curvature is positive and significant, as was noted by Bliss (1997). Finally, macroeconomists have known that the movements of the term structure are well captured by the relationships among level, slope, and curvature; in this respect, the inclusion of macro variables improves predictability, although not correlated with level, slope, and curvature (see Cochrane

Table 6. Posterior estimates for the dynamics of the state equation, namely \mathbf{G} and \mathbf{D}

Variable	Dep. variable: Level		Dep. variable: Curvature	
	Coefficient	Std. error	Coefficient	Std. error
$Level_{t-1}$	0.9984	0.0045	-0.0353	0.0083
$Slope_{t-1}$	0.0330	0.0880	0.9173	0.0163
$Curvature_{t-1}$	0.0031	0.0063	0.0400	0.0120
$Decay_{t-1}$	-0.0010	0.0091	0.1070	0.0330
IP_{t-1}	-0.0009	0.0005	0.0013	0.0010
PCE_{t-1}	0.0003	0.0005	0.0015	0.0009
MI_{t-1}	0.0041	0.0022	0.0040	0.0040
$DEBT_{t-1}$	-0.0030	0.0011	0.0013	0.0020
Constant	0.1222	0.0576	-0.0455	0.1063
Dep. variable: Slope				
$Level_{t-1}$	-0.0353	0.0083	-0.0037	0.00575
$Slope_{t-1}$	0.9173	0.0163	0.0158	0.01128
$Curvature_{t-1}$	0.0400	0.0120	0.0136	0.00827
$Decay_{t-1}$	0.1070	0.0330	0.8944	0.02280
IP_{t-1}	0.0013	0.0010	-0.0005	0.00070
PCE_{t-1}	0.0015	0.0009	0.0002	0.00064
MI_{t-1}	0.0040	0.0041	0.0035	0.00284
$DEBT_{t-1}$	0.0013	0.0021	-0.0034	0.00144
Constant	-0.0455	0.1064	0.1183	0.07351

NOTE: The results are obtained by Monte Carlo averaging, using 10,000 particles of the algorithm.

and Piazzesi 2005 for details). This explains the large posterior standard errors in Table 6.

6.5 Classification of Regime Shifts From the PPM Model

To assess whether any given month is a change point, we compute Bayes factors for each month in our particle algorithm. The Bayes factor, BF , in favor of the hypothesis of a regime change (H_1) against the null hypothesis of no regime change (H_0) is simply the ratio of the posterior odds to the prior odds of the two hypotheses.

$$BF = \frac{P(\text{Change Point at time } t | \text{Data})}{P(\text{No Change Point at time } t | \text{Data})}$$

$$= \frac{P(H_1 | \text{Data})}{P(H_0 | \text{Data})} = \frac{P(Y_t | u_t = 1) P(H_1)}{P(Y_t | u_t = 0) P(H_0)},$$

where $P(Y_t | u_t = i)$ is the averaged likelihood under equation (1) along the MCMC chains at time t if $u_t = i$, with $i = 0, 1$. We initialize by setting $P(H_1)/P(H_0) = 1$.

The monthly time series of Bayes factors should identify regime shifts in the term structure. However, economic shocks to the system, such as a recession, are seldom felt instantaneously. Indeed, there is considerable disagreement as to when the current recession actually started. Thus, our monthly Bayes factors will not always be large enough (say, 2.5 or greater) through all of the months during which the recession (or other shocks) adversely or positively affect the economy: however, if the model is accurate, it should identify at least one or two months in unusual periods in the business cycle with a Bayes factor of at least 3, implying compelling evidence in favor of a regime change.

Figure 6 shows the Bayes factors juxtaposed with the relevant recession periods that were recorded by, among others, Barro and Ursúa (2009) and the Federal Reserve of New York website.³ For the moment, ignore the dashed curve in the graph. Our model correctly classifies all of the months in the term structure that were documented retrospectively as turbulent events. Bayes factors for all of these time points are sufficiently large (at least 3), suggesting a regime shift. However, in the months preceding or succeeding these “shock” months, the Bayes factors hover between 2.5 and 3. This is consistent with our position that economic turbulence in the yield data is not a one-shot event, but rather more of a cumulative effect, culminating in an upward or downward spike in the time series; see also Table 3. In addition, for the recession periods the variability in the Bayes factors depends on the severity and duration of the recession. Thus in the time period 1979–1984, the Bayes factors have the largest variability coinciding with the extreme movements in the actual yield curves. Just as volatility of returns is critical in financial pricing models, the variability of the Bayes factors within and across regime changes provides insight into the magnitude and duration of level and slope shocks to the entire term structure of bond yields. In the next section, we use this concept to show that the Bayes factors serve as leading indicators of economic activity.

In Table 7, our Bayes factors capture 30 out of 81 months in which there were actual macro events that coincided with

Table 7. Characterization of regime changes with Bayes factor ($BF > 2$)

Year	Month	BF	Recession	Inflation	Market	Monetary
1970	3	7.5	X			
1972	9	6.8				X
1972	10	3.3				X
1973	3	2.5		X		X
1973	5	2.3		X		
1973	9	3.0		X		
1973	11	2.3		X		
1974	3	4.6	X	X		
1974	6	2.5	X			
1977	1	2.5				X
1977	10	2.9				X
1980	1	12.0	X			X
1980	5	3.3	X			X
1980	8	2.3	X			X
1980	11	5.5	X	X		X
1981	1	3.9	X	X		X
1981	2	2.3	X	X		X
1981	4	6.8	X	X		X
1982	3	2.5	X	X		X
1984	9	2.9		X	X	X
1985	9	2.5			X	
1990	8	2.8	X			
1995	2	2.9				
1996	9	2.9			X	
1997	6	2.8			X	
1997	8	2.5			X	X
1997	11	2.5			X	X
2000	3	2.3	X			
2000	6	2.3	X			X

NOTE: The actual macroeconomic events, namely the recessions, inflation shocks and market shocks, were recorded by Barro and Ursúa (2009). Monetary shocks are reported at <http://www.newyorkfed.org/markets/statistics/dlyrates/fedrate.html>.

changes to the term structure. These 30 months included 16 months in which the yield curve exhibited moderate to dramatic changes, shown in bold type in the table. It should be emphasized that this is a new empirical result compared with other models in this context that fix the regime changes up front; see Table 1 for competing models. One comparison that we make here is to the work of Chib and Kang (2009), who used a fixed number of change points. They used quarterly data on 16 yields from 1972 to 2007, whereas we use monthly data on 18 yields from 1970 to 2000. They used ATSM, whereas we use the N–S model. Table 7 confirms the five change points of Chib and Kang. In addition, our PPM model also clearly identifies other critical regime changes in the 1970s (due to an oil crisis and inflationary pressures), 1980–1981 period, and 1995–2000.

6.6 Term Structure and Economic Activity

A vast literature originating in the late 1980s documents the empirical regularity with which the slope of the yield curve could serve as a reliable predictor of future real economic activity. The difference between long-term and short-term interest rates (“the slope of the yield curve” or “the term spread”) has demonstrated a consistent negative relationship with subsequent real economic activity in the United States, with a lead time of approximately four to six quarters. The most commonly

³ <http://www.newyorkfed.org/markets/statistics/dlyrates/fedrate.html>.

Table 8. BF vs. CFNAI: VAR(3) estimation and granger causality test

Vector autoregression of order 3, Bayes factor vs. CFNAI							
First equation, Dep. var. <i>BF</i>				Second equation, Dep. var. <i>CF</i>			
R-sq.	0.8356	Rbar-sq.	0.8322	R-sq.	0.9273	Rbar-sq.	0.9261
var(ε_{t1})	0.0135	Q-stat	2.0073	var(ε_{t2})	0.0225	Q-stat	1.4497
Variable	Coeff.	<i>t</i> -stat.	<i>t</i> -prob.	Variable	Coeff.	<i>t</i> -stat.	<i>t</i> -prob.
BF_{t-1}	1.1193	19.009	<0.0001*	BF_{t-1}	-0.1710	-2.367	0.0184*
BF_{t-2}	0.0125	0.1588	0.8738	BF_{t-2}	0.2648	1.8605	0.0636
BF_{t-3}	-0.2298	-4.7910	<0.0001*	BF_{t-3}	0.0636	0.6327	0.5272
CF_{t-1}	-0.0250	-0.5708	0.5684	CF_{t-1}	1.3532	27.579	<0.0001*
CF_{t-2}	0.0596	0.7835	0.4337	CF_{t-2}	-0.3780	-4.232	<0.0001*
CF_{t-3}	-0.0231	-0.3025	0.7624	CF_{t-3}	-0.1484	-2.929	0.0036*
μ_1	0.1243	2.9669	0.0032*	μ_2	0.0260	0.3339	0.7385
Granger causality test			Granger causality test				
Variable	<i>F</i> -value	Prob.	Variable	<i>F</i> -value	Prob.		
<i>BF</i>	756.64	<0.0001*	<i>BF</i>	3.24874	0.0199*		
<i>CF</i>	1.12287	0.3356	<i>CF</i>	1523.26	<0.0001*		

NOTE: R-sq: R^2 coefficient, Rbar-sq: Adjusted R^2 coefficient, Q-stat: Ljung-Box test statistics for residuals autocorrelation, * indicates significance at the 5% level.

used measures of the yield curve are based on differences between interest rates on Treasury securities of contrasting maturities, for instance, 10 years minus 3 months. The measures of real activity for which predictive power has been found include GNP and GDP growth, consumption growth, investment and industrial production, and economic recessions, as recorded by Barro and Ursua (2009). Estrella and Hardouvelis (1991) proposed a simple but effective rule of thumb to identify recessions, namely that the difference between 10-year and 3-month Treasury rates turns negative in advance of recessions. This held true with negative values observed before both the 1990–1991 and 2001 recessions.

Based on the analysis in the preceding section, we now consider the Bayes factors at the change points in the term structure as a measure of change in the yield curve. The Bayes factors are an implicit measure of a sudden change in the intercept, slope, or curvature in the N–S model, because of the state-space evolution; the posterior curves in Figure 4 demonstrate this explicitly. We would like to test whether the Bayes factors can be considered a leading indicator over the business cycle. Our analysis is not meant to be exhaustive, but rather aims to propose a critical application of the output from our analysis to problems that are relevant to macroeconomists.

As a proxy for the business cycle, we use the Chicago Fed National Activity Index (CFNAI).⁴ CFNAI is a monthly index designed to gauge overall economic activity and inflationary pressure, comprising a weighted average of 85 existing monthly indicators of national economic activity. It is constructed to have an average value of 0 and a standard deviation of 1. Because economic activity tends toward trend growth rates over time, a positive index reading corresponds to growth above trend, and a negative index corresponds to growth below trend. The CFNAI also corresponds to the index of economic activity developed by Stock and Watson (1999). Researchers

have found that the CFNAI provides a useful gauge on current and future economic activity and inflation in the United States. In addition, the consensus is that the CFNAI is not a leading indicator of real GDP growth, but rather a coincident one.

Figure 6 depicts the agreement between our Bayes factors and the CFNAI index. The agreement is most pronounced during turbulent periods. These are the critical ones, because they correspond to major regime changes in the term structure.

The foregoing findings lead to the following question of interest: Could the monthly Bayes factor series act as a leading indicator of CFNAI? The direction of causality that this question suggests could prove of value to economists and financial agents because, as noted earlier, this could lead policy makers to better control economic factors, such as Fed Fund rates, mortgage rates, and so on. From a modeling perspective, because the CFNAI is an indicator with “sign,” we take its absolute value to make it comparable to our Bayes factors. A simple third-order vector autoregression to account for quarterly structure in the data is estimated. Denoting the absolute value of the CFNAI, *CF*, and the Bayes Factor, *BF*,

$$\begin{cases} BF_t = \mu_1 + \beta_{11}BF_{t-1} + \beta_{21}BF_{t-2} + \beta_{31}BF_{t-3} \\ \quad + \beta_{41}CF_{t-1} + \beta_{51}CF_{t-2} + \beta_{61}CF_{t-3} + \varepsilon_{t1}, \\ CF_t = \mu_2 + \beta_{12}BF_{t-1} + \beta_{22}BF_{t-2} + \beta_{32}BF_{t-3} \\ \quad + \beta_{42}CF_{t-1} + \beta_{52}CF_{t-2} + \beta_{62}CF_{t-3} + \varepsilon_{t2}. \end{cases}$$

The Granger causality test is predictive. In our context, if the variable BF Granger causes CF, then past values of BF should contain information that helps predict CF above and beyond the information contained in past values of CF alone. From Table 8, when the response variable is BF, none of the CF lags in that equation are significant, whereas when the response variable is CF, the first and second lags of BF are significant, as are the lagged CF values. The overall Granger causal *F* test in the table, with BF as the dependent variable, shows a *p*-value of 0.3356 which is not significant for the collection of regression coefficients corresponding to the lagged CF variables. With CF as the

⁴ http://www.chicagofed.org/economic_research_and_data/cfnai.cfm.

dependent variable, the overall F test for the significance of all the lagged BF variables is highly significant (p -value 0.0199).

Taken together, the empirical facts demonstrate that the Bayes factors from our study could be a leading indicator of economic activity. A reason for this causality stems from the fact that the macro factors enter the dynamic N–S model independently through the observation equation and recursively through the state equation, as developed in Sections 2 and 3. This in turn helps identify regime changes in the entire term structure through the product partitioning of the yield data. But, as we argued earlier, a shock to the economic system is not a one-shot event, and the residual effects of, say, a recession carry over to subsequent months, as confirmed by the analysis in the preceding section. Thus the time series of Bayes factors implicitly contains information that serves as a proxy for a leading economic indicator. This is a completely new empirical finding using PPM-based term structure models, with implications extending beyond term structures.

7. DISCUSSION

The present study contributes to statistical modeling of the term structure in several ways. First, we show that regime changes in yield data can be identified using product partition models along with a dynamic nonlinear N–S model. A comparison with actual events that defined yield regime changes coincided with our model’s classification of these events as regime shifts. Second, we implemented our analysis using particle learning, which allows for faster estimation of the model parameters. Third, the Bayes factors from our analysis could serve as a leading indicator of economic activity, as confirmed via a Granger causality test.

There are several potential extensions to our framework, starting with robust modeling of the error terms with fatter tails and a stochastic volatility effect. However, this would increase the complexity of the algorithm, leading to difficulties in developing prior distributions. Second, the entire specification in this article could be applied to the class of ATSM. Third, it is possible to develop computational strategies to estimate the fixed parameters with backward smoothing without relying on MCMC.

APPENDIX A: BAYESIAN REDUCED-FORM VAR

Jeffreys Prior for \mathbf{E}

The Jeffreys prior for \mathbf{E} takes the form $p(\mathbf{E}) \propto |\mathbf{E}|^{-(M+1)/2}$. The conditional posterior is $\mathcal{IW}(t + M - 2, \hat{\mathbf{E}})$ with $\hat{\mathbf{E}} = (Y - \mathbf{F}\theta_t)'(Y - \mathbf{F}\theta_t)$; Zellner (1971).

The VAR Model

A simple reduced-form VAR model can be written in multiple ways, with each serving a purpose in the development of prior distributions detailed below. Let T denote the length of the time series, k the size of lags, and M the dimension of the vector θ . Consider the following three versions of the VAR model appropriate to this article:

1. $\theta_t = \mathbf{G}\theta_{t-1} + \mathbf{D}\mathbf{K}_{t-1} + w_t$ with $w_t \sim N(0, \mathbf{W})$
2. $\theta_t = \mathbf{X}_t\mathbf{G}\mathbf{D} + w_t$, with $\mathbf{X}_t = [\theta_{t-1} \ K_{t-1}]$ and $\mathbf{G}\mathbf{D} = [G \ D]$, and
3. $\theta_t = \mathbf{Z}_t\mathbf{g}\mathbf{d} + w_t$ with $\mathbf{Z}_t = (I \otimes X_t)$, and $\mathbf{g}\mathbf{d} = \text{vec}(\mathbf{G}\mathbf{D})$.

Let $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_T]$. The OLS estimates are $\hat{\mathbf{g}}\mathbf{d} = (\sum Z_t'Z_t)^{-1} \times (\sum Z_t'\theta_t)$, $\hat{\mathbf{G}}\mathbf{D} = (X'X)^{-1}(X'\theta)$.

Define the sum of squared errors of the VAR by $\hat{S} = (\theta - \mathbf{X}\hat{\mathbf{G}}\mathbf{D})' \times (\theta - \mathbf{X}\hat{\mathbf{G}}\mathbf{D})$, and define the OLS estimate for W as $\hat{W} = \frac{\hat{S}}{\nu}$, with $\nu = T - k$.

The Minnesota Prior

Despite the attractiveness of drawing on cross-sectional information from related economic variables, the VAR model has empirical limitations. For example, in our model the four dependent variables and three lagged independent variables produce a total of 42 coefficients to estimate. Large samples of observations involving time series variables that cover many years are needed to estimate the VAR model, and these are not always available. To overcome these problems, Doan, Litterman, and Sims (1984) proposed using the Minnesota conjugate prior. The intuition underlying this prior is to selectively restrict the hyperparameters of $\mathbf{g}\mathbf{d}$. Toward this end, the prior means and variances take the following form:

$$\mathbf{g}\mathbf{d}_i \sim N(1, \sigma_{\mathbf{g}\mathbf{d}_i}^2),$$

$$\mathbf{g}\mathbf{d}_j \sim N(0, \sigma_{\mathbf{g}\mathbf{d}_j}^2),$$

where $\mathbf{g}\mathbf{d}_i$ denotes the coefficients associated with the lagged dependent variable in each equation of the VAR and $\mathbf{g}\mathbf{d}_j$ represents any other coefficient. The prior means for lagged dependent variables are set to unity in the belief that these are important explanatory variables. On the other hand, a prior mean of 0 is assigned to all other coefficients in the equation. This is like saying that a priori, we believe that the dependent variable should behave as simple random walk, given that the wide recognition in the econometric literature that the random walk performs well in predictions. Let $\mathbf{g}\mathbf{d}$ denote the vector of the $\mathbf{g}\mathbf{d}_i$ and $\mathbf{g}\mathbf{d}_j$ prior means.

The prior variances, $\sigma_{\mathbf{g}\mathbf{d}_i}^2$ and $\sigma_{\mathbf{g}\mathbf{d}_j}^2$, specify uncertainty about the prior mean. Because the VAR model contains a large number of parameters, Doan, Litterman, and Sims (1984) suggested a formula for generating the standard deviations as a function of a small number of hyperparameters, τ and ϕ , and a weighting matrix $z(i, j)$. This approach allows a practitioner to specify individual prior variances for a large number of coefficients in the model using only these three parameters. Doan et al. showed that specification of the standard deviation of the prior is imposed on variable j in equation i at lag k by

$$\sigma_{i,jk} = \tau z(i, j)k^{-\phi} \left(\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}} \right),$$

where $\hat{\sigma}_{ui}$ is the estimated standard error from a univariate autoregression involving only variable i , so that $(\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}})$ is a scaling factor that adjusts for varying magnitudes of the variables across equations i and j . Doan et al. labeled the parameter τ as “overall tightness,” reflecting the standard deviation of the prior on the first lag of the dependent variable. The term $k^{-\phi}$ is a lag decay function with $0 \leq \phi \leq 1$, reflecting the decay rate, a shrinkage of the standard deviation with increasing lag length. This has the effect of forcing the prior means to be zero as the lag length increases, based on the belief that more distant lags represent less important variables in the model. The function $z(i, j)$ specifies the tightness of the prior for variable j in equation i relative to the tightness of the own lags of variable i in equation i .

In this article, the overall tightness and lag decay hyperparameters used in the standard Minnesota prior have values $\tau = 0.1$ and $\phi = 1$. The following weighting matrix is used:

$$Z = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}.$$

This weighting matrix favors $\mathbf{g}\mathbf{d}_i = 1$, because the lagged dependent variable in each equation is considered an important variable. The weighting matrix also enforces a prior mean of 0 for coefficients on other variables in each equation more tightly, because the $\mathbf{g}\mathbf{d}_j$ coefficients are associated with variables considered less important in the

model. Now let $\Sigma_{\mathbf{gd}}$ denote the derived matrix populated by the $\sigma_{i,jk}$ elements as described above.

With the foregoing prior, estimating the VAR follows the same direction as in a standard natural conjugate case. The Gibbs sampler for $\mathbf{W}, \mathbf{G}, \mathbf{D}$ is implemented by sampling from the conditionals $\mathbf{W}|\theta, \mathbf{gd} \sim \mathcal{IW}(v, \hat{S})$, as in the Jeffreys prior case, and $\mathbf{gd}|\mathbf{W}, \theta \sim \mathcal{N}(\mathbf{gd}, \Sigma_{\mathbf{gd}})$, where

$$\overline{\Sigma_{\mathbf{gd}}} = [\Sigma_{\mathbf{gd}}^{-1} + (\mathbf{W}^{-1} \otimes (\mathbf{X}'\mathbf{X}))]^{-1},$$

$$\overline{\mathbf{gd}} = \overline{\Sigma_{\mathbf{gd}}}[\overline{\Sigma_{\mathbf{gd}}}^{-1}\tilde{\mathbf{gd}} + (\mathbf{W}^{-1} \otimes \mathbf{X})'\theta].$$

APPENDIX B: COMPUTATIONAL NOTES AND PRIOR SPECIFICATION

For each pass L of the algorithm (meaning a complete update between time 1 and time T), we use $N = 10,000$ particles, which selects the most likely partition. We repeat the algorithm for $L = 30,000$ passes, for a total of $3E^{10}$ updates for the partitions.

The priors, initializations, and estimation method are given in Table B.1.

Table B.1. Priors, initializations, and estimation method

Parameters	Prior/initialization	Estimation method
$\theta_t = \begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \\ \lambda_t \end{bmatrix}$	$\theta_t^* \sim \mathcal{MVN}(m_t^*, C_t^*)$	$PL^{(+)}$
$\pi(u_t p)$	$p \sim \text{Beta}(\alpha_p, \beta_p)$	PL
\mathbf{E}	Jeffreys	Gibbs
\mathbf{G}, \mathbf{D}	Minnesota	Gibbs
\mathbf{W}	Jeffreys	Gibbs

Descriptions of Model Parameters, Their Priors and Estimation Methods

The second column presents the different prior settings. The θ_t^* are obtained at each time t via a cross-sectional nonlinear maximum likelihood with bootstrapped covariance. *Jeffreys* indicates Jeffreys prior, and *Minnesota* indicates a Minnesota prior with tightness 0.1, decay 1, and weighting matrix as detailed in Appendix A. *PL* indicates Particle Learning, and *Gibbs* indicates standard Gibbs updating. Also as denoted by +, the λ_t coefficient is blindly propagated.

[Received November 2009. Revised February 2011.]

REFERENCES

- Ang, A., and Bekaert, G. (2002), "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20 (2), 163–182. [428]
- Ang, A., and Piazzesi, M. (2003), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics With Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50 (4), 745–787. [428,431]
- Bansal, R., and Zhou, H. (2002), "Term Structure of Interest Rates With Regime Shifts," *Journal of Finance*, 57, 1997–2043. [428]
- Barro, R., and Ursua, J. (2009), "Stock-Market Crashes and Depressions," Working Paper 14760, NBER. [436,437]
- Barry, D., and Hartigan, J. A. (1992), "Product Partition Models for Change Point Problems," *The Annals of Statistics*, 20, 260–279. [427,429]
- (1993), "A Bayesian Analysis for Change Point Problems," *Journal of the American Statistical Association*, 88, 309–319. [427,429]
- BIS (2005), "Zero-Coupon Yield Curves," technical documentation, Bank for International Settlements, Basle. [427]
- Bliss, R. (1997), "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, 9, 197–231. [431,432,435]
- Carvalho, C., Johannes, M., Lopes, H., and Polson, N. G. (2010), "Particle Learning and Smoothing," *Statistical Science*, 25, 88–106. [429]
- Chib, S., and Ergashev, B. (2009), "Analysis of Multifactor Affine Yield Curve Models," *Journal of the American Statistical Association*, 104, 1324–1337. [427]
- Chib, S., and Kang, K. H. (2009), "Change Points in Affine Term-Structure Models: Pricing, Estimation and Forecasting," working paper, Washington University in St. Louis. [427,428,436]
- Christensen, J., Diebold, F. X., and Rudebusch, G. D. (2007), "The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models," Working Paper 2007-20, FRB of San Francisco. [428]
- Cochrane, J., and Piazzesi, M. (2005), "Bond Risk Premia," *American Economic Review*, 95 (1), 138–160. [436]
- Dai, Q., and Singleton, K. (2007), "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," *Review of Financial Studies*, 20 (5), 1669–1706. [428]
- De Pooter, M., Ravazzolo, F., and van Dijk, D. (2007), "Predicting the Term Structure of Interest Rates," discussion paper, Tinbergen Institute, The Netherlands. [431]
- Diebold, F. X., and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337–364. [427,428,430-432,435]
- Diebold, F. X., Li, C., and Yue, V. (2008), "Global Yield Curve Dynamics and Interactions: A Generalized Nelson–Siegel Approach," *Journal of Econometrics*, 146, 351–363. [428,432]
- 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. [427,428,431]
- Doan, T., Litterman, R., and Sims, C. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3 (1), 1–100. [438]
- Duffee, G. R. (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443. [427,435]
- (2006), "Term Structure Estimation Without Using Latent Factors," *Journal of Financial Economics*, 79, 507–536. [435]
- Estrella, A., and Hardouvelis, G. A. (1991), "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46 (2), 555–576. [437]
- Fama, E. F. (2006), "The Behavior of Interest Rates," *The Review of Financial Studies*, 19 (2), 359–379. [427]
- Fama, E. F., and Bliss, R. (1987), "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692. [431]
- Gelfand, A. E., and Ghosh, S. K. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika*, 85, 1–11. [433]
- Koopman, S. J., Mallee, M. I. P., and van der Wel, M. (2010), "Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson–Siegel Model With Time-Varying Parameters," *Journal of Business and Economic Statistics*, 28, 329–343. [427,428]
- Loschi, R. H., and Cruz, F. R. B. (2005), "Extensions to the Product Partition Model: Computing the Probability of a Change," *Computational Statistics and Data Analysis*, 48, 255–268. [429]
- Mumtaz, H., and Surico, P. (2009), "Time-Varying Yield Curve Dynamics and Monetary Policy," *Journal of Applied Econometrics*, 24 (6), 895–913. [428]
- Nelson, C. R., and Siegel, A. F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 473–489. [427]
- Sims, C. A., and Zha, T. (1998), "Bayesian Methods for Dynamic Multivariate Linear Models," *International Economic Review*, 39 (4), 949–968. [427]
- (2006), "Were There Regime Switches in U.S. Monetary Policy?" *The American Economic Review*, 96 (1), 54–81. [427,428]
- Stock, J. H., and Watson, M. W. (1999), "Business Cycle Fluctuations in U.S. Macroeconomic Time Series," in *Handbook of Macroeconomics* (1st ed.), eds. J. B. Taylor and M. Woodford, Vol. 1, USA: Elsevier, Chap. 1, pp. 3–64. [437]
- West, M., and Harrison, P. J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer. [430]
- Yao, Y. C. (1984), "Estimation of a Noisy Discrete-Time Step Function: Bayes and Empirical Bayes Approaches," *The Annals of Statistics*, 12, 1117–1123. [429]
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley. [438]
- Zhu, X., and Rahman, S. (2009), "A Regime-Switching Macro-Finance Model of the Term Structure," working paper, Nanyang Technological University. [428]