Nelson-Siegel models with regime shifts: Is the deterministic adjustment term in the arbitrage-free models important?[1]

Dara Sim
Graduate School of Economics
Osaka University

Masamitsu Ohnishi
Graduate School of Economics & Center for the Study of Finance and Insurance
Osaka University

1 Introduction

The dynamic Nelson-Siegel (DNS) model has drawn attention of many practitioners and academics because of its tractability, and empirical in-sample fitting and out-of-sample forecasting performance, as well as the Nelson-Siegel property. In many studies, it is extended to account for regime shifts in the behavior of term structure dynamics, and is shown to have better forecasting performance than most forecasting models (Xiang & Zhu, 2013). The arbitrage-free version of this, arbitrage-free Nelson-Siegel (ANFS) model differs only by a deterministic adjustment term. In a single-regime setting, Sim & Ohnishi (2014) found that the DNS which does not have this deterministic arbitrage adjustment term appears to outperform the AFNS that does in terms of forecasting Japanese Government Bond (JGB) yields, suggesting that the arbitrage-free condition is not important for forecasting JGB yields. In this study, we investigate the difference between these models within regime-switching setting for the JGB yields data. In particular, we explore within two-regime setting the importance of this deterministic arbitrage adjustment term in terms of fitting and forecasting yield curves, regime identification.

The interest rates are known to exhibit pattern of regime shits. Regime-shift behavior in the interest rates dynamic is attributed to not only monetary policy regimes[2], but also to the changes in business cycle conditions. In order to capture such behavior, regime-shift term structure models are considered. Such models incorporate such important features as flexibility for fitting, ability to capture nonlinear term structure dynamics, ability to mimic violations of the expectation hypothesis, good forecasting performance, etc.

There are many approaches to term structure modeling with regime shifts incorporated. One is to consider the discrete-time affine arbitrage-free term structure models. This kind of models are considered for example in Bansal & Zhou (2002), and Dai et al. (2007). Bansal & Zhou (2002) employ discrete-time CIR models. Assuming constant transition probabilities, and that regime-shift risk is not priced, they find that regime-shift models can account for violations of the expectation hypothesis, and that regimes are closely related to business cycles. On the hand, Dai et al. (2007) empirically implement the discrete-time Gaussian models assuming that regime-shift risk is priced and that transition probabilities are factor-dependent. Another approach is to consider the continuous-time affine arbitrage-free term structure models. This includes Landen (2000), Wu & Zeng (2007), Elliott & Siu (2009). However, due to the model complexity, the empirical implementation based on such models is not well documented. The other approach is the econometric modeling approach. Within this approach, the models are not derived from the arbitrage-free condition. The focus is on fitting and forecasting as well as the interaction between interest rates and other macroeconomic variables. To take this approach, thus it is

---

[2] In order to drive economic growth, central banks of many important developed country such as Japan and the US have adopted ZIRP as part of their monetary policy.

not possible to gain any understanding about market prices of risk. Regime-switching dynamic Nelson-Siegel models considered in Zantedeschi et al. (2011), and Xiang & Zhu (2013) belong to this class of models. Xiang & Zhu (2013) show using US Treasury data that regime-shift DNS has better forecasting performance than most forecasting models.

For empirical study on JGB, Koeda (2013) constructs an arbitrage-free term structure model which subjects short-term rates to regime shifts driven by monetary policy (zero interest rate policy (ZIRP) regime and normal regime). The model is intended to capture features such as the zero lower bound of the policy interest rate and the Bank of Japan's forward guidance of interest rates using the setup of Oda & Ueda (2007) for ZIRP exit rules. He shows that under the ZIRP regime, the effect of deflation (inflation) on lowering (raising) bond yields amplifies on the long end of yield curves, compared with a case under the normal regime.

This study extends the DNS and AFNS models to the case when parameters are subjected to regime shifts. Just like the regime-shift DNS (RSDNS), the regime-shift AFNS (RSAFNS) considered is not derived from the arbitrage-free condition, and so is not an arbitrage-free term structure model. It is just RSDNS with a deterministic term, which is also subjected to regime shifts because the regime-shift volatility parameters appear in the term. As mentioned above since RSDNS is well known to be a good forecasting model, we are interested in the question: Is the term that appears in the single-regime AFNS could be of importance for improving fitting and forecasting yields as well as identifying the regime? To answer this question, this paper conducts empirical study using weekly data of JGB zero-coupon yields.

Estimating the models using Markov Chain Monte Carlo method, we find that RSAFNS outperforms RSDNS for both fitting and forecasting JGB yields. RSAFNS is also able to identify high and low volatility regimes, which is intimately related to the stage of business cycle, much clearer than RSDNS. One drawback of introducing the deterministic term to the RSDNS is that the model's dynamic factors may no longer represent long-term rates, and the slope of the yield curve.

Section 2 describes the RSDNS and RSAFNS models. Section 3 presents the empirical study. Especially, the data used is described, and the regime identification is discussed in detail. The performance comparison between single-regime and regime-shift models are discussed in section 4. Section 5 concludes the paper.

2 Regime-shift DNS and regime-shift AFNS

The regime-shift process is modeled as a discrete-time homogeneous Markov chain S. The regime variable S_t take the value 1 if the state of the world is in regime 1 and 2 if it is in regime 2. Let $\mathbf{P}$ denote the regime transition probability matrix (under physical probability measure), ie., $\mathbf{P}_{ji} = \Pr(S_{t+1} = j | S_t = i)$. The Markov chain S can be represented by a vector process $\boldsymbol{\xi}$ where the random vector $\boldsymbol{\xi}_t$ takes value in the set of canonical unit vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$. It is easy to see that

$$\mathbb{E}[\boldsymbol{\xi}_{t+1} | \boldsymbol{\xi}_t] = \mathbf{P}\boldsymbol{\xi}_t. \tag{1}$$

With this notation, any parameter ψ that depends on regime can be represented as $\psi_{S_t} = \psi(\boldsymbol{\xi}_t) = \boldsymbol{\psi}'\boldsymbol{\xi}_t$, where $\boldsymbol{\psi} = (\psi_0, \psi_1)'$.

Also assume that agents in the economy can observe the current regime and price bonds based on models derived within the non-regime-shift setting, but with the parameters switched according to the current regime. In other words, the regime-shift term structure models considered in this paper are not derived from the arbitrage-free condition. Many papers study the importance of arbitrage-free condition, but there is no consensus among the findings. When

focusing on the forecasting ability of a model, whether it satisfies arbitrage-free condition or not may not be important.

In addition, unlike in Koeda (2013), the models considered in this paper do not subject short-term rates to regime shifts driven by monetary policy. In other words, the models are allowed to identify the regimes, which can be either the ZIRP regime and the normal regime, or something else (such as high and low volatility regimes). We consider in particular the DNS, and AFNS models with regime-shift parameters. They shall be called RSDNS, and RSAFNS respectively.

RSDNS

$$y_t(\tau_n) = \hat{\mathbf{B}}(\tau_n)'\mathbf{X}_t + \epsilon_{n,t}, n = 1, 2, ..., N, \tag{2}$$

$$\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{n,t})'|_{\boldsymbol{\xi}_t} \sim N(0, \sigma_\epsilon^2(\boldsymbol{\xi}_t)\mathbf{I}),$$

where

$$\hat{\mathbf{B}}(\tau_n)' = \begin{bmatrix} 1, & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n}, & \left[\frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} - e^{-\lambda\tau_n}\right] \end{bmatrix} \tag{3}$$

and $y_t(\tau_n)$ is observed zero-coupon yield of time-to-maturity τ_n. Conditional on $\tilde{\boldsymbol{\xi}}_T = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_T)$, $\epsilon_t, t = 1, ..., T$ are assumed serially independent. $\mathbf{X}_t$ is a 3-dimensional vector of latent dynamic factors and is given by

$$\mathbf{X}_t = \boldsymbol{\mu}(\boldsymbol{\xi}_t) + \mathbf{F}\mathbf{X}_{t-1} + \boldsymbol{\nu}_t, \boldsymbol{\nu}_t|_{\boldsymbol{\xi}_t} \sim N(0, \mathbf{G}(\boldsymbol{\xi}_t)) \tag{4}$$

where

$$\boldsymbol{\mu}(\boldsymbol{\xi}_t) = \begin{bmatrix} (1 - e^{-\kappa_L\Delta t})\theta_L(\boldsymbol{\xi}_t) \\ (1 - e^{-\kappa_S\Delta t})\theta_S(\boldsymbol{\xi}_t) \\ (1 - e^{-\kappa_C\Delta t})\theta_C(\boldsymbol{\xi}_t) \end{bmatrix}, \mathbf{F} = \begin{bmatrix} e^{-\kappa_L\Delta t} & 0 & 0 \\ 0 & e^{-\kappa_S\Delta t} & 0 \\ 0 & 0 & e^{-\kappa_C\Delta t} \end{bmatrix}, \tag{5}$$

and

$$\mathbf{G}(\boldsymbol{\xi}_t) = \begin{bmatrix} \frac{\sigma_L^2(\boldsymbol{\xi}_t)}{2\kappa_L}\left(1 - e^{-2\kappa_L\Delta t}\right) & 0 & 0 \\ 0 & \frac{\sigma_S^2(\boldsymbol{\xi}_t)}{2\kappa_S}\left(1 - e^{-2\kappa_S\Delta t}\right) & 0 \\ 0 & 0 & \frac{\sigma_C^2(\boldsymbol{\xi}_t)}{2\kappa_C}\left(1 - e^{-2\kappa_C\Delta t}\right) \end{bmatrix}. \tag{6}$$

Conditional on $\tilde{\boldsymbol{\xi}}_T = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_T)$ $\boldsymbol{\nu}_t, t = 1, ..., T$ are assumed serially independent, and are independent of $\epsilon_t, t = 1, ..., T$.

RSAFNS

$$y_t(\tau_n) = \hat{\mathbf{B}}(\tau_n)'\mathbf{X}_t + \hat{A}(\tau_n, \boldsymbol{\xi}_t) + \epsilon_{n,t}, n = 1, 2, ..., N, \tag{7}$$

$$\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{n,t})'|_{\boldsymbol{\xi}_t} \sim N(0, \sigma_\epsilon^2(\boldsymbol{\xi}_t)\mathbf{I}),$$

where

$$\hat{\mathbf{B}}(\tau_n)' = \begin{bmatrix} 1, & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n}, & \left[\frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} - e^{-\lambda\tau_n}\right] \end{bmatrix} \tag{8}$$

$$\hat{A}(\tau_n, \boldsymbol{\xi}_t) = -\sigma_L^2(\boldsymbol{\xi}_t)\frac{\tau_n^2}{6} - \sigma_S^2(\boldsymbol{\xi}_t)\left[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda\tau_n}}{\tau_n} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda\tau_n}}{\tau_n}\right]$$

$$-\sigma_C^2(\boldsymbol{\xi}_t)\left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2} - \frac{1}{4\lambda}\tau_n e^{-2\lambda\tau_n} - \frac{3}{4\lambda^2}e^{-2\lambda\tau_n} - \frac{2}{\lambda^3}(1 - e^{-\lambda\tau_n}) + \frac{5}{8\lambda^2}\frac{1-e^{-2\lambda\tau_n}}{\tau_n}\right] \tag{9}$$

and $y_t(\tau_n)$ is observed zero-coupon yield of time-to-maturity τ_n. $\mathbf{X}_t$ is a 3-dimensional vector of latent dynamic factors and is given as in (4). The assumptions about the conditional independence of ϵ_t and $\boldsymbol{\nu}_t$ are the same as that of RSDNS. Conditional on $\tilde{\boldsymbol{\xi}}_T$ and $\tilde{\boldsymbol{X}}_T = (\boldsymbol{X}_1, ..., \boldsymbol{X}_T)$, $\boldsymbol{\nu}_t, t = 1, .., T$ are assumed serially independent, and independent of $\epsilon_t, t = 1, ..., T$.

Note that each element of $\mathbf{X}_t = (X_{L,t}, X_{S,t}, X_{C,t})$ is corresponding to level factor, slope factor, and curvature factor. The parameters $\theta_i, \kappa_i, \sigma_i, i \in \{L, S, C\}$ respectively represent the factors' long run means, mean reversion rates, and volatilities. λ is the decay rate parameter that determines the speed of decay of loadings on slope and curvature. In the above models, except for the mean reversion rates and decay rate parameter, the long run means and volatilities are subjected to regime shifts.

3 Empirical study

3.1 Data

We use JGB zero-coupon yields constructed by Kikuchi & Shintani (2012) using the the method proposed by Steeley (1991). Kikuchi & Shintani (2012) show that the Steeley's method is the best among popular yield-curve estimation methods based on some considered criteria. The daily data based on this method is downloaded from Bank of Japan (BoJ)'s website. We convert daily data to weekly data by choosing estimates on Thursday (if available, and on Wednesday if not) as the weekly estimates. We skip those weeks where both the estimates on Thursday and Wednesday are not available.[3] The time-to-maturity spectrum included in our study are 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 15, 18, and 20 year, and the sample consists of 692 weekly observations covering the period from January 1999 to December 2012.

Figure 1 plots 6-month, 5, 10, and 20-year yields over the sample periods considered. It is easy to see from this figure that the short rate shifts between two regimes: the near-zero rate regime, and the normal regime. In fact, BoJ introduced and lifted zero-interest rate policy (ZIRP) several times according to the macroeconomic conditions. During the sample period considered, ZIRP was lifted for the period from August 2000 to February 2001, and for the period from July 2006 to the end of 2008. Figure 2 shows the result of fitting a Markov mixed normal distribution model to the 6-month yield. In the ZIRP regime ($S_t = 1$), $y(6m) \sim N(0.0022, 0.0006^2)$, with the transition probabilities: $Pr(S_t = 1|S_{t-1} = 1) = 0.9947$, and $Pr(S_t = 2|S_{t-1} = 1) = 0.0053$. On the other hand, in the normal regime ($S_t = 2$), $y(6m) \sim N(0.0060, 0.0011^2)$, with the transition probabilities: $Pr(S_t = 2|S_{t-1} = 2) = 0.9870$, and $Pr(S_t = 1|S_{t-1} = 2) = 0.0130$. The expected period of staying in the ZIRP regime (resp. normal regime) is about 3.6 years (resp. 1.5 years).

3.2 Estimation method

The models are estimated using the Markov Chain Monte Carlo (MCMC) method. To generate factor and regime variables, the forward and backward sampling approach is employed. As initial values for the regime variable, the realizations drawn from smoothed probabilities obtained from fitting Markov mixed normal distribution model to the 6-month yield as described above are used. All linear parameters can be generated by Gibbs sampling algorithm. The decay rate parameter λ in both models, and the factor volatility parameters $\sigma_L^2, \sigma_S^2, \sigma_C^2$, and mean reversion rates $\kappa_L, \kappa_S, \kappa_C$ in the RSAFNS model, are nonlinear. These parameters are generated from the Random Walk Metropolis (RWM) algorithm. In the RWM step, calibrating scale parameter is crucial to achieve good approximation to the target distribution. If it is too small, we will

---

[3]Although daily data are available, we adopt weekly data instead for ease of computation, and for better long horizon forecasts.
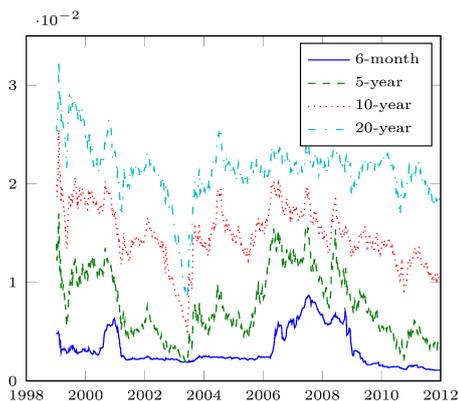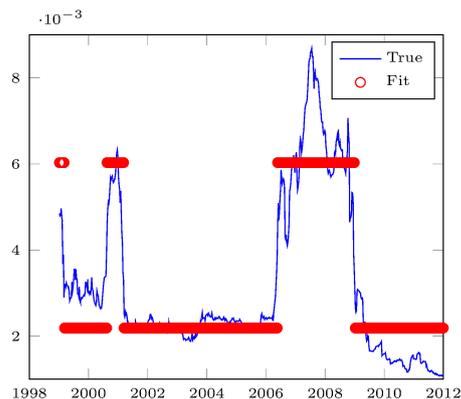
Figure 1: Zero-coupon yields

Figure 2: Fitting MS mixed normal distribution model to 6-month yields.

accept every draws and the convergence is slow. Too high the scale parameter, we will likely reject every draws. Optimally efficient acceptance rate for RWM is known to be 0.234 (Roberts, Gelman, Gilks et al., 1997). To calibrate the scale parameter for RWM, the algorithm used by Feldhütter & Nielsen (2012) is employed. At every 100th draw, the acceptance ratio of each parameter is evaluated. If it is smaller than 0.2, the scale is cut in half, and if it is larger than 0.5, the scale is doubled. This calibration step is carried out for the first half the burnt-in period because the convergence result only applies if the scale is constant. For both models, 25,000 draws are generated. Out of this, 10,000 draws are used as the burnt-in period. See Appendix A for detail derivation of posterior functions.

3.3 Estimation result

In this subsection, the estimation results of both models are presented. Especially, the focus is on the regime identification and factor interpretation.

3.3.1 Results for RSDNS

Table 1 reports parameter estimates and standard errors (given in the parenthesis), as well p-value for the null hypothesis that parameter estimates are equal under both regimes. Figure 3a shows area plot of smoothed probabilities[4] along with line plots of a short-term rate (0.5-year yield), a slope (difference between 10-year and 0.5-year yields), and a long-term rate (20-year yield). Figure 3b shows area plot of smoothed probabilities of regime 1 along with line plots of the estimated level factor X_L, the negative of slope factor $-X_S$, and the estimated short rate $r = X_L + X_S$. First, the estimates of volatilities under both regimes are not significantly different. However, the absolute value of the slope's long-run mean θ_S in regime 1 is significant smaller than that in regime 2 at 10% significant level. Thus, regime 1 is identified as regime of flattening yield curves. From Figure 3a it easy to confirm that regime 1 identified by the model corresponds to the period of decreasing in the slope of yield curves because long-term rates are falling and/or short term-rates are rising.

There can be three possible situations that lead to the flattening of yield curves. First, long-term rates shrink because market expects that future inflation is falling. Second, flattening yield curves indicate anticipation of future economic downturn. Third, as part of monetary

---

[4]The smoothed probabilities are obtained using smoother algorithm described in Kim & Nelson (1999).

policy when inflation and unemployment rates reach a target level, BoJ lifts ZIRP allowing short-term rates to rise even if long-term rates are decreasing. In contrast to regime 1, regime 2 can be identified as period of steepening yield curves characterized by increasing long-term rates and slope, which in turn indicates market expectations for future economic growth and rising inflation.

Table 1: Estimation result (RSDNS)

	Regime 1	Regime 2	p-value
θ_L	0.0233 (0.0048)*	0.0313 (0.0026)*	0.2306
θ_S	−0.0178 (0.0050)*	−0.0295 (0.0026)*	0.0714
θ_C	−0.0378 (0.0296)	−0.0338 (0.0167)*	0.8760
σ_L	0.0040 (0.0003)*	0.0041 (0.0002)*	0.4429
σ_S	0.0045 (0.0003)*	0.0046 (0.0002)*	0.7055
σ_C	0.0121 (0.0008)*	0.0131 (0.0007)*	0.3310
σ_e	0.0003 (0.0000)*	0.0006 (0.0000)*	0.0000
κ_L	0.7062 (0.2689)*	−	−
κ_S	0.7917 (0.2856)*	−	−
κ_C	0.6242 (0.3328)	−	−
λ	0.3542 (0.0014)*	−	−
$Pr(S_t = i\|S_{t-1} = i)$	0.9674 (0.0075)*	0.9856 (0.0043)*	−

Note:* indicates the estimate is significantly different from 0 at 5% significance level. The standard errors are given in the parenthesis. p-value is for the null hypothesis that parameter estimates are equal under both regimes.

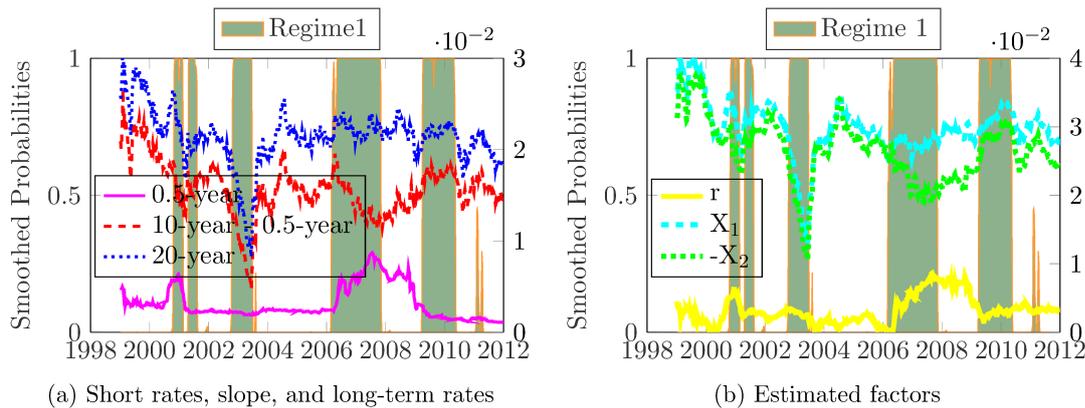(a) Short rates, slope, and long-term rates   (b) Estimated factors

Figure 3: Smoothed probabilities of regime 1 (RSDNS)

Next, we are interested in the interpretation of the estimated paths of the dynamic factors. Table 3 reports correlations between the estimated path of the level, slope, curvature, and short rate, and the empirical level, slope, curvature, and short-term rates, respectively. Three rows of correlation values are reported for each factor. The first row shows values for the whole sample period. The second shows values for the subsample period from the 1st week of January 1999 to the 3rd week of July 2006, whereas the third shows those for the subsample period from the last week of July 2006 to the last week of December 2011. For the slope and curvature factors, the correlation with their empirical counterparts is high for the 3 sample periods. For the level factor, the correlation with the long-term rates is weak for the whole sample period and high for subsample period 1. However, for subsample period 2, there is almost no correlation at all.

It may appear that the NS-type models fail to maintain their propert for period of economic turmoil.

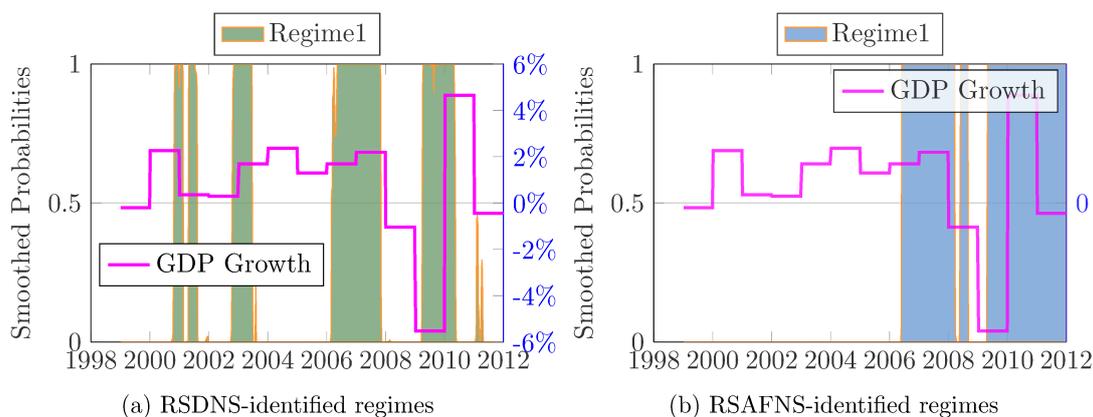(a) RSDNS-identified regimes   (b) RSAFNS-identified regimes

Figure 4: Smoothed probabilities of regime 1 and GDP growth

In addition to regime identification and factor interpretation, estimates of other parameters are important too. First, it is a stylized fact that interest rate factors are highly persistent. The estimated mean reversion rates are so small indicating that the three factors are highly persistent. Also, the estimates of the decay parameter λ is 0.3542, which is about the same as the estimate from the DNS model. Furthermore, from the estimated value of the transition probabilities, we are able to calculate an expected period of stay in each regime, which is about 0.59 years for regime 1, and about 1.34 years for regime 2. Finally, since the observed yields tend to be higher in regime 2 than in regime 1, so are the observed errors. Indeed, the estimated standard deviation is 6bps in regime 2 compared to 3bps in regime 1.

3.3.2 Results for RSAFNS

The estimation results of the RSAFNS model are given in Table 2, Figure 4b, 5, and 6. First, consider what we can say about the two regimes identified by the model. The estimated values of factor volatilities are significantly higher in regime 2 than in regime 1, and so regime 2 is characterized by high volatilities, and regime 1 is characterized by low volatilities. It is widely documented that high volatility is often observed during economic recession (Dai et al. (2007), Xiang & Zhu (2013)). Indeed, this is the case for our result. From Figure 4b, it is obvious that the high volatility regime consists of a prolonged period of slow economic growth, and the 2008-2009 period of economic recession. This result is even more obvious than that of the RSDNS model (where the estimates of factor volatilities are not significantly different). On the other hand, the low volatility regime is characterized by the 2006-2007 period of flattening yield curves, and the year 2010 period of economic recovery. It is also interesting to note that the 2006-2007 period of flattening yield curves is followed by the 2008-2009 period of economic recession (or period of steepening yield curves), which is in turn followed by the year 2010 period of economic recovery. Once again, this suggests that the economic condition was as anticipated by the market. Thus, RSAFNS identifies the two regimes as the high volatility regime (bad time regime) with large magnitude of deterministic adjustment term (see Figure 6), and the low volatility regime (normal regime) with small magnitude of deterministic adjustment term. The expected period of stay in high (resp. low) volatility regime is 3.21 (resp. 1.13) years.

In addition, we investigate if the dynamic factors can be interpreted as long-term rates, slope, and curvature of yield curves. For the three sample periods considered, the level factor

Table 2: Estimation result (RSAFNS)

	Regime 1	Regime 2	p-value
θ_L	0.0324 (0.0031)*	0.0374 (0.0050)*	0.4015
θ_S	-0.0239 (0.0043)*	-0.0363 (0.0042)*	0.0359
θ_C	-0.0528 (0.0089)*	-0.0337 (0.0074)*	0.1264
σ_L	0.0020 (0.0002)*	0.0077 (0.0001)*	0.0000
σ_S	0.0031 (0.0003)*	0.0055 (0.0003)*	0.0000
σ_C	0.0115 (0.0008)*	*0.0169 (0.0013)	0.0001
σ_e	0.0003 (0.0000)*	0.0005 (0.0000)*	0.0000
κ_L	0.6418 (0.3998)	$-$	
κ_S	0.5315 (0.2982)	$-$	
κ_C	0.5599 (0.2555)*	$-$	
λ	0.2881 (0.0002)*	$-$	
$Pr(S_t = i | S_{t-1} = i)$	0.9830 (0.0055)*	0.9940 (0.0026)*	

Note:* indicates the estimate is significantly different from 0 at 5% significance level. The standard errors are given in the parenthesis. p-value is for the null hypothesis that parameter estimates are equal under both regimes.

(a) Short rates, slope, and long-term rates
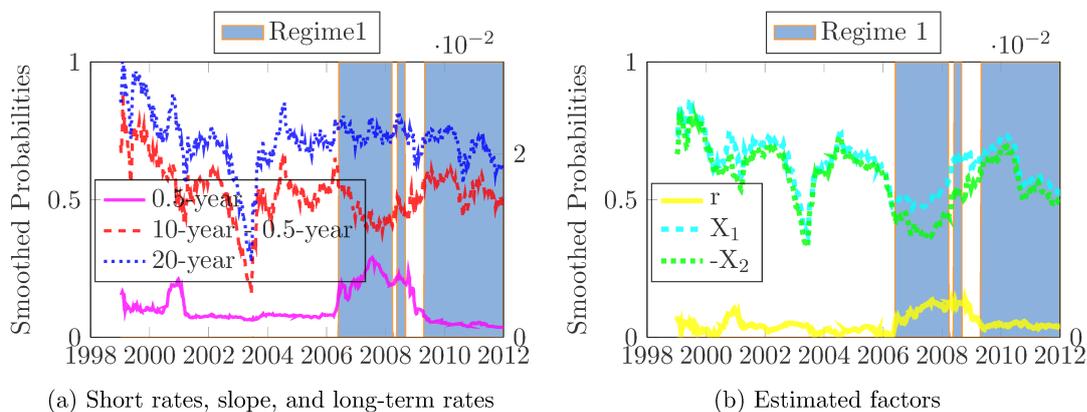
(b) Estimated factors

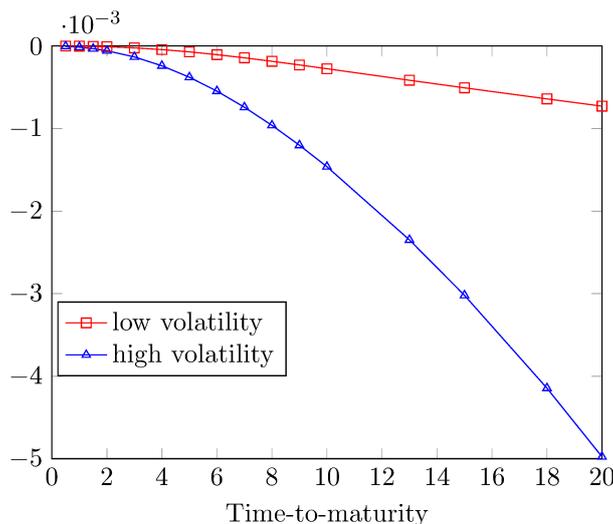Figure 5: Smoothed probabilities of regime 1 (RSAFNS)

Figure 6: RSAFNS's adjustment term

and slope factor have weak correlation with their empirical counterparts. The reason for this is that the estimate of the decay rate parameter λ is only about 0.29 much lower than 0.35, which mean that the loadings on the slope and curvature factors decay slower than in the RSDNS model. Finally, notice that for the subsample period 2, the level factor appears to have no correlation with the long-term rates at all. Once again, this suggests that the NS-type models fail to maintain the NS property for the period of economic turmoil.

3.3.3 The estimated paths of the short rate

It is well known that affine term structure models that do not ensure positivity of the short rate are not good at dealing with situation where the instantaneous interest rates are near zero lower bound (ZLB). Figure 7 shows plots of the paths of the short rate estimated by the single-regime models. As can be seen, when the interest rates are near ZLB, the filtered estimates of the short rate are likely to take on negative value. Now we want to investigate whether allowing parameters in DNS and AFNS to switch between regimes can improve the filtered estimates of the short rate. Figure 8 show plots of the paths of the short rate estimated by RSDNS and RSAFNS. It is obvious that the regime-shift models can better capture short rate near ZLB than the single-regime models. Furthermore, the filtered estimates of the short rate implied by RSAFNS take on almost non-negative value at all.

4   Performance comparison

This section focuses on the comparison, in term of in-sample fitting and out-of-sample forecasting, of five competing models: DNS, AFNS, MAFNS, RSDNS, and RSAFNS.

4.1   Fitting performance

RMSEs of the five models are plotted in Figure 9. It is not obvious to tell which one among these have the best performance. However, the regime-shift models are able to fit better for the short and long end of the yield curves, for which the single regime models appear to have difficulty. The average RMSEs of DNS, AFNS, MAFNS, RSDNS, and RSAFNS, are respectively 4.83,

Table 3: Identified factors

	13-year	15-year	18-year	20-year
X_1^{RSDNS}	0.7467	0.8087	0.8512	0.8679
	0.8690	0.8959	0.9184	0.9321
	−0.1452	0.0105	0.2469	0.2881
X_1^{RSAFNS}	0.5610	0.5975	0.6132	0.6373
	0.7270	0.7652	0.7954	0.8174
	−0.1345	0.0268	0.2145	0.1714

	0.5-year − 13-year	0.5-year −15-year	0.5-year − 18-year	0.5-year − 20-year
X_2^{RSDNS}	0.9169	0.9338	0.9396	0.9543
	0.9279	0.9451	0.9643	0.9730
	0.8529	0.9407	0.9752	0.9781
X_2^{RSAFNS}	0.7630	0.7718	0.7690	0.7920
	0.7970	0.8296	0.8606	0.8817
	0.6492	0.7917	0.8524	0.8246

	2*5-year −0.5- − 13-year	2*5-year −0.5- −15-year	2*5-year −0.5- − 18-year	2*5-year −0.5- −20-year
X_3^{RSDNS}	0.9391	0.9686	0.9784	0.9681
	0.9543	0.9709	0.9704	0.9584
	0.9829	0.9892	0.9933	0.9915
X_3^{RSAFNS}	0.9263	0.9351	0.9273	0.9206
	0.9635	0.9754	0.9755	0.9638
	0.8692	0.8927	0.9026	0.8892

	0.5-year	1-year	1.5-year	2-y
r^{RSDNS}	0.7747	0.7603	0.6953	0.6274
	0.6160	0.4016	0.2414	0.1273
	0.9156	0.8677	0.8416	0.8173
r^{RSAFNS}	0.8489	0.8263	0.7674	0.7086
	0.6906	0.5116	0.3596	0.2468
	0.9383	0.9058	0.8870	0.8692

Note: The values reported are the correlation between the estimated level, slope, curvature factors, short rate (rows), and the empirical level, slope, curvature, and short-term rates (columns), respectively. iy denotes i-y yields. All sample (1st row): 08-Jan-1999 to 30-Dec-2011. Subsample1 (2nd row): 08-Jan-1999 to 21-Jul-2006. Subsample2 (3rd row): 28-Jul-2006 to 30-Dec-2011.
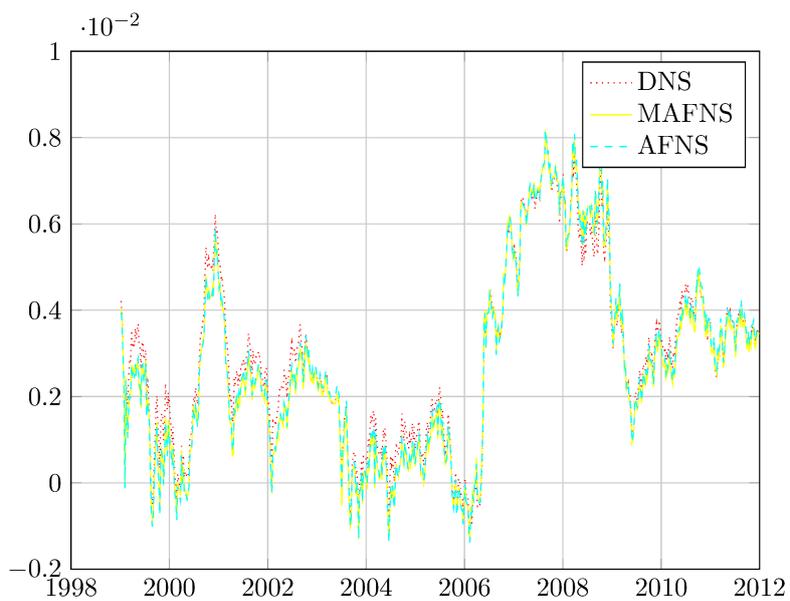
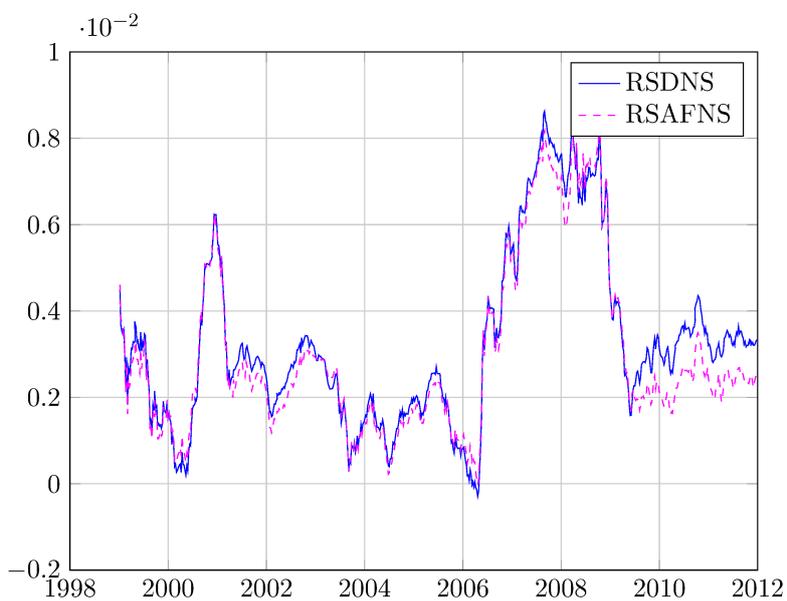Figure 7: Estimated paths of short rates (DNS, AFNS, MAFNS)

Figure 8: Estimated paths of short rates (RSDNS, RSAFNS)

4.59, 458, 4.76, 4.41bps. It seems that over-all RSAFNS has the best fit. Nevertheless, since the three single-regime models already have very good fitting performance, in terms of this the advantage of introducing regime-shift parameters is slim.
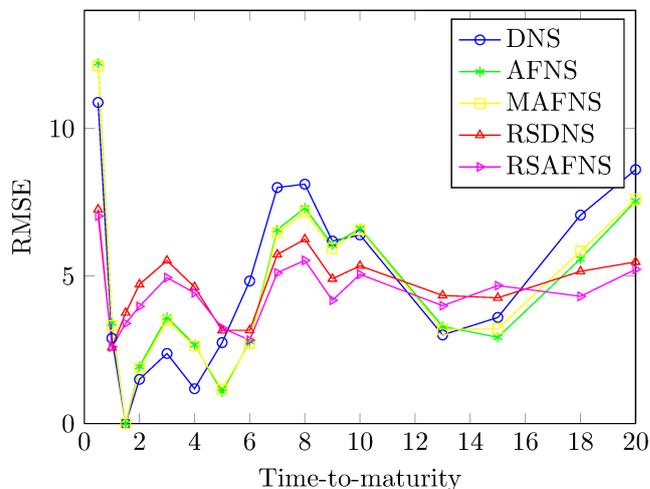

Figure 9: Root mean square errors

4.2 Forecasting performance

To compare out-of-sample forecasting performance, consider forecasting horizons of 4 weeks and 12 weeks. Forecasts are made for the period from the 2nd week of November 2010 to the last week of December 2011. We obtain 58 observations for the 4-week-ahead forecasts, and 50 observations for the 12-week-ahead forecasts (Appendix B shows how to compute forecasts and forecast errors' covariance matrix). Table 4 reports the value of RMSEFs in basis points, as well as the value of Diebold-Mariano (DM) statistic. For the 4-week-ahead forecasts, RSAFNS outperforms RSDNS for almost all maturities. For the 12-week forecasts, except for maturities less than 5 years RSAFNS outperform RSDNS for most of the maturities. It is thus almost obvious that RSAFNS has the best overall performance.

Figure 10 shows plots of RMSFEs for the three single-regime models and the two regime-shift models. Clearly, the regime-shift models dominate the single-regime models for almost all the maturities and for both forecasting horizons. The reason for this can probably be attributed to the fact that the regime-shift models are able to capture dynamics of short rate at the near ZLB better than the single-regime models.

4.3 Economic value of predicting excess return

In this section, the economic value of the excess-return predictability of RSAFNS in compare to RSDNS is evaluated. Consider a mean-variance investor who constructs every four weeks a dynamically rebalanced portfolio of N risky zero-coupon bonds and one risk-free zero-coupon bond. Her problem is to minimize the portfolio variance subject to a given target portfolio expected excess return xr_p^*, i.e.,

$$\min_{\boldsymbol{w}_t} \left\{ \boldsymbol{w}_t' \Sigma_{t+h|t}^{\mathrm{rx}} \boldsymbol{w}_t \right\}$$
$$\text{s.t. } \boldsymbol{w}_t' \boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}} = \mathrm{xr}_p^*$$

Table 4: Forecast results

Maturity	4-week-ahead			12-week-ahead		
	RSDNS	RSAFNS	DM	RSDNS	RSAFNS	DM
0.5	8.348	3.985	−4.480	8.6751	11.014	3.980
1	1.457	2.768	3.297	2.2885	6.699	3.664
1.5	4.142	1.061	−4.604	3.130	4.990	1.052
2	6.112	2.910	−5.686	4.740	5.141	0.204
3	6.989	0.922	−3.056	6.226	7.129	0.420
4	6.650	6.020	−1.198	7.540	8.817	0.708
5	6.630	6.909	1.051	10.664	11.960	1.132
6	9.044	3.406	−2.341	14.148	14.147	−0.001
7	11.068	3.976	−3.498	16.924	15.562	−1.203
8	10.489	4.505	−3.282	16.541	14.303	−1.845
9	9.128	3.902	−2.812	14.891	12.429	−2.136
10	8.416	0.989	−2.236	13.885	11.353	−1.995
13	7.864	7.611	−0.400	12.562	10.594	−1.622
15	7.046	6.990	−0.140	10.981	9.052	−2.269
18	7.801	7.314	−1.154	10.280	8.773	−1.727
20	7.608	7.955	1.042	11.968	11.003	−1.519

This table reports RMSFE (in basis points) and Diebold Mariano (DM) statistics. Under the null hypothesis of equal forecasting accuracy, the DM statistic follows a standard normal distribution.
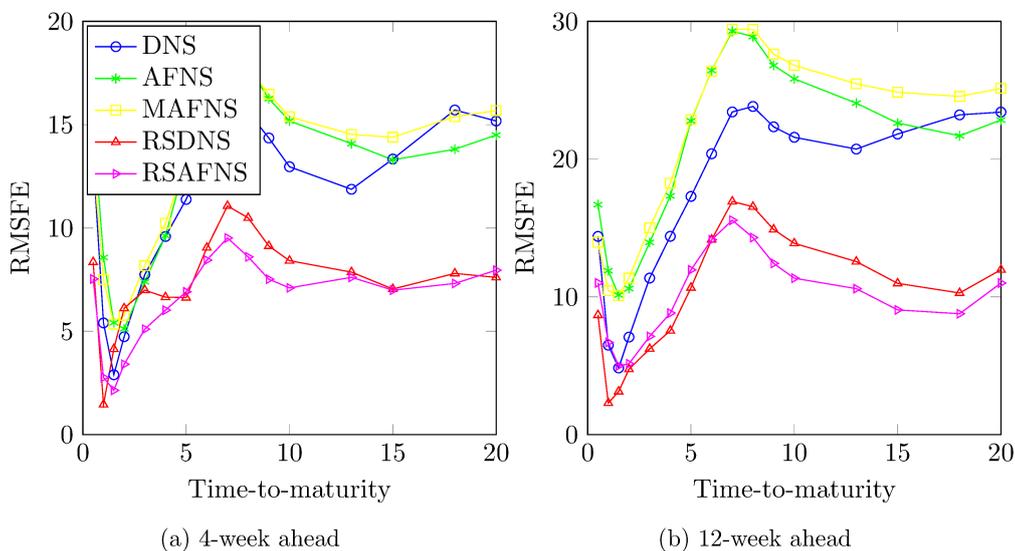

(a) 4-week ahead   (b) 12-week ahead

Figure 10: Root mean square forecast errors

where w_t is a $(N, 1)$-vector of portfolio weights on the risky zero-coupon bonds; $\mu^{rx}_{t+h|t}$, and $\Sigma^{rx}_{t+h|t}$ are respectively the conditional expectation, and variance of bond excess return (Appendix C shows how these values can be computed). Solving this optimization problem give us

$$w_t = \frac{xr^*_p}{C_t}(\Sigma^{rx}_{t+h|t})^{-1}\mu^{rx}_{t+h|t}, \tag{10}$$

where $C_t = (\mu^{rx}_{t+h|t})'(\Sigma^{rx}_{t+h|t})^{-1}\mu^{rx}_{t+h|t}$. If $\mu^{rx}_{t+h|t}$, and $\Sigma^{rx}_{t+h|t}$ are calculated from a model $\mathcal{M}$, then the optimal weights are those given by $\mathcal{M}$. Denote by $w^{\mathcal{M}}_t$ these optimal weights. The portfolio gross return of this strategy is

$$R^{\mathcal{M}}_{p,t+h} = 1 + r_f T_h + (w^{\mathcal{M}}_t)'rx_{t+h},$$

where rx_{t+h} is the excess returns of the risky zero-coupon bonds at time $t + h$, T_h is the holding period.

To assess the economic value of the predictability of RSAFNS, the performance fee as in Della Corte et al. (2008), Thornthon & Valente (2012), and Xiang & Zhu (2013) is adopted. A performance fee is defined as a maximum fee a quadratic utility investor would be willing to pay to switch from the benchmark model (RSDNS) to the competitive model (RSAFNS). That is the performance fee Φ is such that

$$\sum_{j=t_f}^{N_i-h}\left\{\left(R^{RSAFNS}_{p,t+h} - \Phi\right) - \frac{\delta}{2(1+\delta)}\left(R^{RSAFNS}_{p,t+h} - \Phi\right)^2\right\} = \sum_{j=t_f}^{N_i-h}\left\{R^{RSDNS}_{p,t+h} - \frac{\delta}{2(1+\delta)}(R^{RSDNS}_{p,t+h})^2\right\}, \tag{11}$$

where δ denotes the investor's degree of relative risk aversion, N_i is the number of periods of investment. It is easy to see that the performance fee defined as in (11) is positive means that RSAFNS has positive economic value compared to RSDNS. Since, RSAFNS nests RSDNS, the improvement in the forecasting ability can be totally attributed the deterministic term. Thus, the positive value of the fee suggests that the deterministic term has an economic value in terms of forecasting bond excess return.

To evaluate the economic value of the deterministic term, this study calculates the performance fee measure for an investment period of seven months. Since the portfolio is rebalanced dynamically every month, the 1-month yield is the risk free rate. However, the data of 1-month yield is not available. It is thus obtained from linear interpolation. Also $2, 3, ..., 20$-y zero-coupon bonds are considered as largest admissible set of risky bonds.

The out-of-sample performance fees are reported in Table 5 for different value of target portfolio excess returns and different admissible set of risky bonds. For example, for the set $\{2,3,...,20\}$, for a target portfolio excess return of 1% (annual rate), the performance fee is 5.2488 basis points. Not surprisingly, the fees increase with the increase in target portfolio excess return. The positivity suggests that the deterministic term has economic value for forecasting the bonds $\{2,3,...,20\}$' excess return. However, with different set of bonds, the performance fee switches sign. It turns out to be negative for the admissible set $\{6,7,...,20\}$. So it is not possible to come to a conclusion using this measure. Nevertheless, when considering different combinations of risky bonds, as reported Table 6, it turns out that as the number of bonds available for investing increases, the positive ratio of the performance fee also increases. This suggests that the strategy given by RSAFNS is more likely to perform better than RSDNS. This in turn suggests an economic value of the deterministic term for forecasting bond excess return.

Table 5: Out-of-sample performance fee

Risky bonds	$\mathrm{xr}_p^* = 1\%$	$\mathrm{xr}_p^* = 1.5\%$	$\mathrm{xr}_p^* = 2\%$	$\mathrm{xr}_p^* = 2.5\%$	$\mathrm{xr}_p^* = 3\%$
{2,3,...,20}	5.2488	7.8719	10.4952	13.1187	15.7424
{2,3,...,10}	12.3551	18.5385	24.7267	30.9124	37.1048
{6,7,...,20}	−1.3964	−2.0955	−2.7942	−3.4926	−4.1906
{6,7,...,19}	4.5972	6.8972	9.1993	11.5034	13.8095

Note: The investor's relative risk aversion is set to $\delta = 5$. The target portfolio excess returns are annual rates in percents, and Φ's are annual rates in basis points.

Table 6: Signs of the performance fees for different combinations of risky bonds ($\delta = 5, \mathrm{xr}_p^* = 2\%$)

k-combination	Positive	Negative	Positive Ratio
2	107	64	0.6257
3	717	268	0.7279
4	3033	843	0.7825
5	9583	2045	0.8241

Note: This table report the number of k-combinations (from the set of 19 risky bonds {2,3,...20}) where Φ is positive, and negative, as well as the ratio of the number of positives.

5 Summary

In this paper, the dynamic Nelson-Siegel model and the arbitrage-free Nelson Siegel model with parameters subjected to regime shifts is considered. Just like RSDNS, RSAFNS is not derived from the arbitrage-free condition, and so is not an arbitrage-free term structure model. RSAFNS is just RSDNS with a deterministic term that is also subjected to regime shifts. Since RSDNS is well known to be a good forecasting model, it may be interesting to examine if the arbitrage-free adjustment term which appears in the single-regime AFNS could be of importance for improving fitting and forecasting yields as well as identifying the regime. The answer to this question is a 'YES'. The empirical study on Japanese Government Bond zero-coupon yields showed that RSAFNS outperformed RSDNS for both fitting and forecasting JGB yields. RSAFNS was also able to identify high and low volatility regimes, which are related to the stage of business cycle, much clearer than RSDNS. One drawback of introducing the deterministic term to RSDNS is that the model's dynamic factors may no longer represent level, slope as in the RSDNS model. In addition, we found that the regime-shift models remarkably dominate the single-regime NS-type models for forecasting JGB yields. Therefore, for a forecasting model, RSAFNS should be preferred to RSDNS, and the single-regime NS-type models.

A MCMC Algorithm

Given the observed zero-coupon yields $\tilde{y}_T := (y_1, ..., y_T)$, we estimate $\tilde{X}_T := (X_1, ..., X_T)$, $\tilde{\xi}_T := (\xi_1, ..., \xi_T)$, and the parameters $\Theta := (\theta', \kappa', \sigma', \lambda, q', \sigma_\epsilon')'$, where $\theta = (\theta_L', \theta_S', \theta_C')'$, $\kappa = (\kappa_L, \kappa_S, \kappa_C)'$, $\sigma = (\sigma_L', \sigma_S', \sigma_C')'$, $q = (q_{11}, q_{22})'$ (with $q_{11} = \Pr(\xi_t = e_1 | \xi_{t-1} = e_1)$, $q_{22} = \Pr(\xi_t = e_2 | \xi_{t-1} = e_2))$, $\sigma_\epsilon = (\sigma_{\epsilon,1}, \sigma_{\epsilon,2})'$.

Note that the generations of $\tilde{X}_T, \tilde{\xi}_T, \lambda, q$, and σ_ϵ are almost identical for both models; however, the generations of θ, κ, σ are different depending on each model. Therefore, I first describe the algorithm for the generations $\tilde{X}_T, \tilde{\xi}_T, \lambda, q$, and σ_ϵ for both models, and then the algorithms for the generations of θ, κ, σ for each model.

First of all, the following conditional posteriors (likelihood functions) are useful.

- Given $\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T$, the conditional posterior of $\tilde{\boldsymbol{y}}_T$ is

$$p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T) = (2\pi)^{-T/2}\prod_{t=1}^{T}(\sigma_e^2 \cdot \boldsymbol{\xi}_t)^{-N/2}\exp\left\{-\frac{1}{2}(\sigma_e^2 \cdot \boldsymbol{\xi}_t)^{-1}\hat{\boldsymbol{\epsilon}}_t'\hat{\boldsymbol{\epsilon}}_t\right\},$$

where $\hat{\boldsymbol{\epsilon}}_t = \boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{X}_t$ for the DNS model, and $\hat{\boldsymbol{\epsilon}}_t = \boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{X}_t - \boldsymbol{A}\cdot\boldsymbol{\xi}_t$ for the AFNS model.

- For $i = L, S, C$, the conditional posterior of $\tilde{X}_{i,T}$ given $\Theta, \tilde{\boldsymbol{\xi}}_T$ is

$$p(\tilde{X}_{i,T}|\Theta, \tilde{\boldsymbol{\xi}}_T) = p(X_{i,1}|\Theta, \tilde{\xi}_1)\prod_{t=2}^{T}(2\pi\sigma_{\nu_i}^2 \cdot \boldsymbol{\xi}_t)^{-1/2}\exp\left\{-\frac{1}{2}(\sigma_{\nu_i}^2 \cdot \boldsymbol{\xi})^{-1}\hat{\nu}_{i,t}^2\right\},$$

where $\hat{\nu}_t = X_{i,t} - e^{-\kappa_i\Delta t}X_{i,t-1} - (1 - e^{-\kappa_i\Delta t})(\boldsymbol{\theta}_i \cdot \boldsymbol{\xi}_t)$, and $\sigma_{\nu_i}^2 = \frac{\sigma_i^2}{2\kappa_i}(1 - e^{-2\kappa_i\Delta t})$.

- Conditional posterior of $\tilde{\boldsymbol{\xi}}_T$ given Θ is

$$p(\tilde{\boldsymbol{\xi}}_T|\Theta) = p(\boldsymbol{\xi}_1)\prod_{t=2}^{T}q_{11}^{1\{\boldsymbol{\xi}_t=\boldsymbol{e}_1,\boldsymbol{\xi}_{t-1}=\boldsymbol{e}_1\}}(1 - q_{11})^{1\{\boldsymbol{\xi}_t=\boldsymbol{e}_2,\boldsymbol{\xi}_{t-1}=\boldsymbol{e}_1\}}q_{22}^{1\{\boldsymbol{\xi}_t=\boldsymbol{e}_2,\boldsymbol{\xi}_{t-1}=\boldsymbol{e}_2\}}$$

$$(1 - q_{22})^{1\{\boldsymbol{\xi}_t=\boldsymbol{e}_1,\boldsymbol{\xi}_{t-1}=\boldsymbol{e}_2\}}$$

where $1_{\{\bullet\}}$ is the indicator function.

Generations of $\tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \lambda, \boldsymbol{q}, \sigma_\epsilon$ for both models

1. Conditional on $\tilde{\boldsymbol{y}}_T, \tilde{\boldsymbol{\xi}}_T$, and Θ, generate $\tilde{\boldsymbol{X}}_T = (\boldsymbol{X}_1, ..., \boldsymbol{X}_T)$ using the forward filtering and backward sampling (FFBS) approach.

2. Conditional on $\tilde{\boldsymbol{y}}_T, \tilde{\boldsymbol{X}}_T$, and Θ, generate $\tilde{\boldsymbol{\xi}}_T = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_T)$ using the FFBS approach.

3. Conditional on $\tilde{\boldsymbol{y}}_T, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T$, and $\Theta_{\backslash\lambda}$, generate λ using Random Walk Metropolis algorithm. Draw λ from

$$\lambda_* = \lambda_m + \varsigma v$$

where $v \sim N(0,1)$, and ς is the scaling factor used to adjust the acceptance rate. The acceptance probability is

$$\alpha = \min\left\{\frac{p(\lambda_*|\Theta_{\backslash\lambda}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}{p(\lambda_m|\Theta_{\backslash\lambda}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}, 1\right\}$$

The conditional posterior of λ is

$$p(\lambda|\Theta_{\backslash\lambda}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T) \propto p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{X}}_T|\Theta, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{\xi}}_T|\Theta)p(\lambda|\Theta_{\backslash\lambda})$$

$$\propto p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T)p(\lambda).$$

Assuming flat prior, we obtain the posterior ratio as

$$\frac{p(\lambda_*|\Theta_{\backslash\lambda}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}{p(\lambda_m|\Theta_{\backslash\lambda}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)} = \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\sigma_e^2 \cdot \boldsymbol{\xi}_t)^{-1}[\hat{\boldsymbol{\epsilon}}_{*,t}'\hat{\boldsymbol{\epsilon}}_{*,t} - \hat{\boldsymbol{\epsilon}}_{m,t}'\hat{\boldsymbol{\epsilon}}_{m,t}]\right\},$$

where $\hat{\boldsymbol{\epsilon}}_{*,t}$, and $\hat{\boldsymbol{\epsilon}}_{m,t}$ are $\hat{\boldsymbol{\epsilon}}_t$ evaluated at λ_*, and λ_m respectively.

4. Conditional on $\tilde{y}_T, \tilde{X}_T, \tilde{\xi}_T$, and $\Theta_{\backslash q}$, generate q_{11}, and q_{22} from beta distributions. The conditional posterior of $q = (q_{11}, q_{22})'$ is

$$p(q|\Theta_{\backslash q}, \tilde{X}_T, \tilde{\xi}_T, \tilde{y}_T) \propto p(\tilde{y}_T|\Theta, \tilde{X}_T, \tilde{\xi}_T)p(\tilde{X}_T|\Theta, \tilde{\xi}_T)p(\tilde{\xi}_T|\Theta)p(q|\Theta_{\backslash q})$$
$$\propto p(\tilde{\xi}_T|q)p(q).$$

The conjugate priors of q_{11}, q_{22} are beta distributions $q_{11} \sim \text{beta}(u_{11}, u_{12}), q_{22} \sim \text{beta}(u_{22}, u_{21})$. That is,

$$p(q) \propto q_{11}^{u_{11}-1}(1 - q_{11})^{u_{12}-1}q_{22}^{u_{22}-1}(1 - q_{22})^{u_{21}-1}.$$

Then,

$$p(q|\Theta_{\backslash q}, \tilde{X}_T, \tilde{\xi}_T, \tilde{y}_T) \propto q_{11}^{n_{11}+u_{11}-1}(1 - q_{11})^{n_{12}+u_{12}-1}q_{22}^{n_{22}+u_{22}-1}(1 - q_{22})^{n_{21}+u_{21}-1}.$$

Thus, the posteriors of q_{11}, q_{22} are the beta distributions

$$q_{11}|_{...} \sim \text{beta}(n_{11} + u_{11}, n_{12} + u_{12}),$$
$$q_{22}|_{...} \sim \text{beta}(n_{22} + u_{22}, n_{21} + u_{21}),$$

where n_{jk} denotes the number of transitions from state j to k, which can be easily counted for given $\tilde{\xi}_T = (\xi_1, ..., \xi_T)$.

5. Conditional on $\tilde{y}_T, \tilde{X}_T, \tilde{\xi}_T$, and $\Theta_{\backslash \sigma_\epsilon^2}$, generate $\sigma_\epsilon^2 = (\sigma_{\epsilon,1}^2, \sigma_{\epsilon,2}^2)'$ from inverted Gamma distributions. The conditional posterior of σ_ϵ^2 is

$$p(\sigma_\epsilon^2|\Theta_{\backslash \sigma_\epsilon^2}, \tilde{X}_T, \tilde{y}_T, \tilde{\xi}_T) \propto p(\tilde{y}_T|\Theta, \tilde{X}_T, \tilde{\xi}_T)p(\tilde{X}_T|\Theta, \tilde{\xi}_T)p(\tilde{\xi}_T|\Theta)p(\sigma_\epsilon^2|\Theta_{\backslash \sigma_\epsilon^2})$$
$$\propto p(\tilde{y}_T|\Theta, \tilde{X}_T, \tilde{\xi}_T)p(\sigma_\epsilon^2).$$

The conjugate prior of $\sigma_{\epsilon,j}^2, j = 1, 2$ is an inverted Gamma distribution $\text{IG}(\frac{v_0}{2}, \frac{\delta_0}{2})$, i.e.,

$$p(\sigma_{\epsilon,j}^2) \propto (\sigma_{\epsilon,j}^2)^{-\frac{v_0}{2}-1} \exp\left\{-\frac{\delta_0}{2}(\sigma_{\epsilon,j}^2)^{-1}\right\}.$$

Thus, the posterior distribution of $\sigma_{\epsilon,j}^2, j = 1, 2$ are the inverted Gamma distributions

$$\sigma_{\epsilon,j}^2|_{...} \sim \text{IG}(\frac{v_j}{2}, \frac{\delta_j}{2}), \quad v_j = v_0 + Nn_j, \delta_j = \delta_0 + \sum_{t \in \mathcal{T}_j} \hat{\epsilon}_t'\hat{\epsilon}_t,$$

where n_j denotes the number of times of being in state j, and $\mathcal{T}_j$ denotes a set of times of being in state j.

Generations of θ, κ, σ for RSDNS

For RSDNS, θ, κ, σ appear only in the transition equations. Let

$$a_i := e^{-\kappa_i \Delta t}$$
$$\mu_i := (1 - e^{-\kappa_i \Delta t})\theta_i$$
$$\sigma_{\nu_i}^2 := \sigma_i^2 \frac{1 - e^{-2\kappa_i \Delta t}}{2\kappa_i}$$

for $i = L, S, C$. The transition equations can now be written as

$$X_{i,t} = \xi_t'\mu_i + a_i X_{i,t-1} + \nu_i, \nu_i \sim N(0, \sigma_{\nu_i}^2 \cdot \xi_t).$$

1. Conditional on $\tilde{X}_T, \tilde{\xi}_T, \sigma^2_{\nu_i}$, generate $(\mu'_i, a_i)', i = L, S, C$ from multivariate normal distributions. For a conjugate multivariate normal prior of $(\mu'_i, a_i)' \sim N(m_{i,0}, \Sigma_{i,0})$, the posterior is

$$\begin{bmatrix} \mu_i \\ a_i \end{bmatrix} \sim N(m_{i,1}, \Sigma_{i,1}), \quad \Sigma_{i,1} = (\Sigma_{i,0}^{-1} + Z_i'Z_i)^{-1}, m_{i,1} = \Sigma_{i,1}(\Sigma_{i,0}^{-1}m_{i,0} + Z_i'Y_i)$$

where

$$Z_i = \begin{bmatrix} z_{i,2} \\ ... \\ z_{i,T} \end{bmatrix}, z_{i,t} = \left[\frac{\xi_t'}{\sqrt{\sigma^2_{\nu_i} \cdot \xi_t}}, \frac{X_{i,t-1}}{\sqrt{\sigma^2_{\nu_i} \cdot \xi_t}} \right],$$

$$Y_i = \left[\frac{X_{i,2}}{\sqrt{\sigma^2_{\nu_i} \cdot \xi_2}}, ..., \frac{X_{i,T}}{\sqrt{\sigma^2_{\nu_i} \cdot \xi_T}} \right]'.$$

2. Conditional on $\tilde{X}_T, \tilde{\xi}_T, \mu, a$, generate $\sigma^2_{\nu_{i,j}}, i = L, S, C, j = 1, 2$ from inverted Gamma distributions. For a conjugate inverted Gamma prior $\sigma^2_{\nu_{i,j}} \sim IG(\frac{v_0}{2}, \frac{\delta_0}{2})$, the posterior is

$$\sigma^2_{\nu_{i,j}}|... \sim IG(\frac{v_{i,j}}{2}, \frac{\delta_{i,j}}{2}), \quad v_{i,j} = v_0 + n_j, \delta_{i,j} = \delta_0 + \sum_{t \in \mathcal{T}_j} \hat{\nu}^2_{i,t}$$

where $\hat{\nu}_{i,t} = X_{i,t} - \xi_t'\mu_i - a_iX_{i,t-1}$.

Generations of θ, κ, σ for RSAFNS

1. Conditional on $\tilde{X}_T, \tilde{\xi}_T, \Theta$, generate θ_i from normal distributions. The conditional posterior of θ_i is

$$\begin{aligned} p(\theta_i|\Theta_{\backslash\theta_i}, \tilde{X}_T, \tilde{\xi}_T, \tilde{y}_T) &\propto p(\tilde{y}_T|\Theta, \tilde{X}_T, \tilde{\xi}_T)p(\tilde{X}_T|\Theta, \tilde{\xi}_T)p(\tilde{\xi}_T|\Theta)p(\theta_i|\Theta_{\backslash\theta_i}) \\ &\propto p(\tilde{X}_{i,T}|\Theta, \tilde{\xi}_T)p(\theta_i). \end{aligned}$$

For a conjugate normal prior $\theta_i \sim N(m_{i,0}, \Sigma_{i,0})$, the posterior is

$$\theta_i|... \sim N(m_{i,1}, \Sigma_{i,1}), \quad \Sigma_{i,1} = (\Sigma_{i,0}^{-1} + Z_i'Z_i)^{-1}, m_{i,1} = \Sigma_{i,1}(\Sigma_{i,0}^{-1}m_{i,0} + Z_i'Y_i)$$

where

$$Z_i = \begin{bmatrix} z_{i,2} \\ ... \\ z_{i,T} \end{bmatrix}, z_{i,t} = \frac{1 - e^{-\kappa_i \Delta t}}{\sqrt{(\sigma^2_i \cdot \xi_t)\frac{1-e^{-2\kappa_i \Delta t}}{2\kappa_i}}}\xi_t',$$

$$Y_i = \left[\frac{X_{i,2} - e^{-\kappa_i \Delta t}X_{i,1}}{\sqrt{(\sigma^2_2 \cdot \xi_2)\frac{1-e^{-2\kappa_i \Delta t}}{2\kappa_i}}}, ..., \frac{X_{i,T} - e^{-\kappa_i \Delta t}X_{i,T-1}}{\sqrt{(\sigma^2_T \cdot \xi_T)\frac{1-e^{-2\kappa_i \Delta t}}{2\kappa_i}}} \right]'.$$

2. Conditional on $\tilde{y}_T, \tilde{X}_T, \tilde{\xi}_T$, and $\Theta_{\backslash\kappa_i}$, generate $\kappa_i, i = L, S, C$ using Random Walk Metropolis algorithm. Draw κ_i from

$$\kappa_i^{(*)} = \kappa_i^{(m)} + \varsigma v,$$

where $v \sim N(0,1)$, and ς is the scaling factor used to adjust the acceptance rate. The acceptance probability is

$$\alpha = \min\left\{\frac{p(\kappa_i^{(*)}|\Theta_{\backslash \kappa_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}{p(\kappa_i^{(m)}|\Theta_{\backslash \kappa_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}, 1\right\}.$$

Since the conditional posterior of κ_i is

$$p(\kappa_i|\Theta_{\backslash \kappa_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T) \propto p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{X}}_T|\Theta, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{\xi}}_T|\Theta)p(\kappa_i|\Theta_{\backslash \kappa_i})$$

$$\propto p(\tilde{X}_{i,T}|\Theta, \tilde{\boldsymbol{\xi}}_T)p(\kappa_i),$$

assuming flat prior, we obtain the posterior ratio as

$$\frac{p(\kappa_i^{(*)}|\Theta_{\backslash \kappa_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}{p(\kappa_i^{(m)}|\Theta_{\backslash \kappa_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)} = \exp\left\{-\frac{1}{2}\sum_{t=2}^{T}\left[log\left(\frac{(1-e^{2\kappa_i^{(*)}\Delta t})\kappa_i^{(m)}}{(1-e^{2\kappa_i^{(m)}\Delta t})\kappa_i^{(*)}}\right)\right.\right.$$

$$\left.\left. + (\sigma_i^2 \cdot \boldsymbol{\xi}_t)^{-1}\left[\frac{2\kappa_i^{(*)}}{1-e^{-2\kappa_i^{(*)}\Delta t}}(\hat{\nu}_{i,t}^{(*)})^2 - \frac{2\kappa_i^{(m)}}{1-e^{-2\kappa_i^{(m)}\Delta t}}(\hat{\nu}_{i,t}^{(m)})^2\right]\right]\right\},$$

where $\hat{\nu}_{i,t}^{(*)}$, and $\hat{\nu}_{i,t}^{(m)}$ are $\hat{\nu}_{i,t}$ evaluated at $\kappa_i^{(*)}$, and $\kappa_i^{(m)}$ respectively.

3. Conditional on $\tilde{\boldsymbol{y}}_T, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T$, and $\Theta_{\backslash \sigma_i}$, generate $\sigma_i, i = L, S, C$ using Random Walk Metropolis algorithm. Draw σ_i from

$$\sigma_i^{(*)} = \sigma_i^{(m)} + \Omega v,$$

where $v \sim N(0, I_2)$, and Ω is the scaling factor used to adjust the acceptance rate. The acceptance probability is

$$\alpha = \min\left\{\frac{p(\sigma_i^{(*)}|\Theta_{\backslash \sigma_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}{p(\sigma_i^{(m)}|\Theta_{\backslash \sigma_i}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T)}, 1\right\},$$

where the conditional posterior of σ_i is

$$p(\sigma_i|\Theta_{\backslash \sigma_i^2}, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T, \tilde{\boldsymbol{y}}_T) \propto p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{X}}_T|\Theta, \tilde{\boldsymbol{\xi}}_T)p(\tilde{\boldsymbol{\xi}}_T|\Theta)p(\sigma_i|\Theta_{\backslash \sigma_i})$$

$$\propto p(\tilde{\boldsymbol{y}}_T|\Theta, \tilde{\boldsymbol{X}}_T, \tilde{\boldsymbol{\xi}}_T)p(\tilde{X}_{i,T}|\Theta, \tilde{\boldsymbol{\xi}}_T)p(\sigma_i)$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{t=2}^{T}\left[log(\sigma_i^2 \cdot \boldsymbol{\xi}_t) + (\sigma_i^2 \cdot \boldsymbol{\xi}_t)^{-1}\frac{2\kappa_i}{1-e^{-2\kappa_i\Delta t}}\hat{\nu}_{i,t}^2\right]\right.$$

$$\left. -\frac{1}{2}\sum_{t=1}^{T}(\sigma_\epsilon^2 \cdot \boldsymbol{\xi}_t)^{-1}\hat{\epsilon}_t'\hat{\epsilon}_t\right\}p(\sigma_i).$$

Noting the errors of the transition equations do not depend on σ_i and assuming flat prior, we obtain the posterior ratio as

$$\frac{p(\sigma_i^{(*)}|...)}{p(\sigma_i^{(m)}|...)} = \exp\left\{-\frac{1}{2}\sum_{t=2}^{T}\left[log\left(\frac{(\sigma_i^{(*)})^2 \cdot \boldsymbol{\xi}_t}{(\sigma_i^{(m)})^2 \cdot \boldsymbol{\xi}_t}\right) + \left[(\sigma_i^{(*)} \cdot \boldsymbol{\xi}_t)^{-2} - (\sigma_i^{(m)} \cdot \boldsymbol{\xi}_t)^{-2}\right]\frac{2\kappa_i}{1-e^{-2\kappa_i\Delta t}}\hat{\nu}_{i,t}^2\right]\right.$$

$$\left. -\frac{1}{2}\sum_{t=1}^{T}(\sigma_\epsilon^2 \cdot \boldsymbol{\xi}_t)^{-1}\left[(\hat{\epsilon}_t^{(*)})'\hat{\epsilon}_t^{(*)} - (\hat{\epsilon}_t^{(m)})'\hat{\epsilon}_t^{(m)}\right]\right\},$$

where $\hat{\epsilon}_t^{(*)}$, and $\hat{\epsilon}_t^{(*)}$ are resp. measurement errors evaluated at $\sigma_i^{(*)}$ and $\sigma_i^{(m)}$.

B Forecasts and forecast errors' covariance matrix

For a h forecasting horizon, the forecast of yields at time t is $\boldsymbol{y}_{t+h|t} = \mathbb{E}[\boldsymbol{y}_{t+h}|\boldsymbol{X}_t, \boldsymbol{\xi}_t]$, and the forecast errors' covariance matrix is $\Sigma_{t+h|t} = \mathbb{E}[(\boldsymbol{y}_{t+h} - \boldsymbol{y}_{t+h|t})(\boldsymbol{y}_{t+h} - \boldsymbol{y}_{t+h|t})'|\boldsymbol{X}_t, \boldsymbol{\xi}_t]$. We can write[5]

$$\begin{aligned}
\boldsymbol{y}_{t+h} &= \boldsymbol{A}\boldsymbol{\xi}_{t+h} + \boldsymbol{B}\boldsymbol{X}_{t+h} + \boldsymbol{\epsilon}_{t+h}, \\
\boldsymbol{X}_{t+h} &= \boldsymbol{\mu}\boldsymbol{\xi}_{t+h} + \boldsymbol{F}\boldsymbol{X}_{t+h-1} + \boldsymbol{\nu}_{t+h},
\end{aligned}$$

as

$$\boldsymbol{X}_{t+h} = \sum_{j=0}^{h-1}\boldsymbol{F}^j\boldsymbol{\mu}\boldsymbol{\xi}_{t+h-j} + \boldsymbol{F}^h\boldsymbol{X}_t + \sum_{j=0}^{h-1}\boldsymbol{F}^j\boldsymbol{\nu}_{t+h-j},$$

$$\boldsymbol{y}_{t+h} = \boldsymbol{A}\boldsymbol{\xi}_{t+h} + \boldsymbol{B}\left(\sum_{j=0}^{h-1}\boldsymbol{F}^j\boldsymbol{\mu}\boldsymbol{\xi}_{t+h-j} + \boldsymbol{F}^h\boldsymbol{X}_t\right) + \sum_{j=0}^{h-1}\boldsymbol{B}\boldsymbol{F}^j\boldsymbol{\nu}_{t+h-j} + \boldsymbol{\epsilon}_{t+h}.$$

Since

$$\boldsymbol{\xi}_{t+h|t} = \mathbb{E}[\boldsymbol{\xi}_{t+h}|\boldsymbol{\xi}_t] = \boldsymbol{P}^h\boldsymbol{\xi}_t,$$

we have

$$\begin{aligned}
\boldsymbol{y}_{t+h|t} &= \boldsymbol{A}\mathbb{E}_t[\boldsymbol{\xi}_{t+h}] + \sum_{j=0}^{h-1}\boldsymbol{B}\boldsymbol{F}^j\boldsymbol{\mu}\mathbb{E}_t[\boldsymbol{\xi}_{t+h-j}] + \boldsymbol{B}\boldsymbol{F}^h\boldsymbol{X}_t \\
&= \boldsymbol{A}\boldsymbol{P}^h\boldsymbol{\xi}_t + \boldsymbol{B}\left(\sum_{j=0}^{h-1}\boldsymbol{F}^j\boldsymbol{\mu}\boldsymbol{P}^{h-j}\boldsymbol{\xi}_t + \boldsymbol{F}^h\boldsymbol{X}_t\right), \quad (12)
\end{aligned}$$

$$\begin{aligned}
\Sigma_{t+h|t} &= \sum_{j=0}^{h-1}\boldsymbol{B}\boldsymbol{F}^j\mathbb{E}_t[\boldsymbol{\nu}_{t+h-j}\boldsymbol{\nu}'_{t+h-j}](\boldsymbol{B}\boldsymbol{F}^j)' + \mathbb{E}_t[\boldsymbol{\epsilon}_{t+h}\boldsymbol{\epsilon}'_{t+h}] \\
&= \sum_{j=0}^{h-1}\boldsymbol{B}\boldsymbol{F}^j\mathbb{E}[\boldsymbol{G}(\boldsymbol{\xi}_{t+h-j})|\boldsymbol{\xi}_t](\boldsymbol{B}\boldsymbol{F}^j)' + \boldsymbol{I}\mathbb{E}[\sigma_\epsilon^2(\boldsymbol{\xi}_{t+h})|\boldsymbol{\xi}_t] \\
&= \sum_{j=0}^{h-1}\boldsymbol{B}\boldsymbol{F}^j\boldsymbol{G}(\boldsymbol{P}^{h-j}\boldsymbol{\xi}_t)(\boldsymbol{B}\boldsymbol{F}^j)' + (\sigma_\epsilon^2 \cdot (\boldsymbol{P}^h\boldsymbol{\xi}_t))\boldsymbol{I}, \quad (13)
\end{aligned}$$

where to obtain (13) the forecast errors of the regime variable is assumed zero (i.e., $\boldsymbol{\xi}_{t+k} - \mathbb{E}_t[\boldsymbol{\xi}_{t+k}] = 0$).

C h-period rebalanced portfolio of zero-coupon bonds

For notational convenient, denote the zero-coupon yield of a time-to-maturity τ by y_t^τ instead of $y_t(\tau)$.

The price at time t of a zero-coupon bond with time-to-maturity τ (in year) is

$$P_t(\tau) = e^{-y_t^\tau \tau}.$$

---

[5]Note that for RSDNS, $\boldsymbol{A}$ is a matrix of zeros.

Consider investing in one unit of τ_n-(zero-coupon) bond. The return after holding for h periods $(h\Delta t = T_h$ years) is

$$log\left(\frac{P_{t+h}(\tau_n - T_h)}{P_t(\tau_n)}\right) = -y_{t+h}^{\tau_n - T_h}(\tau_n - T_h) + y_t^{\tau_n}\tau_n$$

Note that for h periods of holding (frequency of portfolio rebalancing), the return on T_h-bond is a risk free rate $y_t^{T_h} = r_f$ (annual rate). The excess return of the τ_n-bond $\mathrm{rx}_{t+h}^{\tau_n}$ is then

$$\mathrm{rx}_{t+h}^{\tau_n} = -y_{t+h}^{\tau_n - T_h}(\tau_n - T_h) + y_t^{\tau_n}\tau_n - y_t^{T_h}T_h = -y_{t+h}^{\tau_n - T_h}(\tau_n - T_h) + y_t^{\tau_n}\tau_n - r_f T_h.$$

The excess returns of a set of zero-coupon bonds are

$$\mathrm{rx}_{t+h} = -\boldsymbol{y}_{t+h}^{\boldsymbol{\tau} - T_h}\mathbf{1} \odot (\boldsymbol{\tau} - T_h\mathbf{1}) + \boldsymbol{y}_t^{\boldsymbol{\tau}} \odot \boldsymbol{\tau_n} - r_f T_h\mathbf{1},$$

where $\boldsymbol{\tau} = (\tau_1, ..., \tau_N)'$, $\boldsymbol{y}_t^{\boldsymbol{\tau}} = (y_t^{\tau_1}, ..., y_t^{\tau_N})'$, and $\odot$ denotes the element-by-element product. The conditional expectation and variance-covariance matrix are

$$\begin{aligned}
\boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}} &= \mathbb{E}_t[\mathrm{rx}_{t+h}] = -\boldsymbol{y}_{t+h|t}^{\boldsymbol{\tau} - T_h}\mathbf{1} \odot (\boldsymbol{\tau} - T_h\mathbf{1}) + \boldsymbol{y}_t^{\boldsymbol{\tau}} \odot \boldsymbol{\tau_n} - r_f T_h\mathbf{1} \\
&= -\tilde{\boldsymbol{y}}_{t+h|t} + \boldsymbol{y}_t^{\boldsymbol{\tau}} \odot \boldsymbol{\tau_n} - r_f T_h\mathbf{1} \quad (14) \\
\boldsymbol{\Sigma}_{t+h|t}^{\mathrm{rx}} &= \mathbb{E}_t[(\mathrm{rx}_{t+h} - \mathrm{rx}_{t+h|t})(\mathrm{rx}_{t+h} - \mathrm{rx}_{t+h|t})'] = \tilde{\boldsymbol{\Sigma}}_{t+h|t} \quad (15)
\end{aligned}$$

From (12),

$$\tilde{\boldsymbol{y}}_{t+h|t} = \tilde{\boldsymbol{A}}\boldsymbol{P}^h\boldsymbol{\xi}_t + \tilde{\boldsymbol{B}}\left(\sum_{j=0}^{h-1}\boldsymbol{F}^j\boldsymbol{\mu}\boldsymbol{P}^{h-j}\boldsymbol{\xi}_t + \boldsymbol{F}^h\boldsymbol{X}_t\right).$$

From (13),

$$\tilde{\boldsymbol{\Sigma}}_{t+h|t} = \sum_{j=0}^{h-1}\tilde{\boldsymbol{B}}\boldsymbol{F}^j\boldsymbol{G}(\boldsymbol{P}^{h-j}\boldsymbol{\xi}_t)(\tilde{\boldsymbol{B}}\boldsymbol{F}^j)' + (\sigma_\epsilon^2 \cdot (\boldsymbol{P}^h\boldsymbol{\xi}_t))\mathrm{diag}(\boldsymbol{\tau} - T_h\mathbf{1}),$$

where

$$\tilde{\boldsymbol{A}} = \left[\begin{array}{c} \hat{\boldsymbol{A}}(\tau_1 - T_h)'(\tau_1 - T_h) \\ ... \\ \hat{\boldsymbol{A}}(\tau_N - T_h)'(\tau_N - T_h) \end{array}\right], \quad \tilde{\boldsymbol{B}} = \left[\begin{array}{c} \hat{\boldsymbol{B}}(\tau_1 - T_h)'(\tau_1 - T_h) \\ ... \\ \hat{\boldsymbol{B}}(\tau_N - T_h)'(\tau_N - T_h) \end{array}\right].$$

Consider a mean-variance investor who optimally invests in N risky zero-coupon bonds, and one risk-free zero-coupon bond. She constructs monthly dynamically rebalanced portfolio by minimizing portfolio variance subject to a given target portfolio expected return r_p^*. Her problem is

$$\min_{\boldsymbol{w}_t}\left\{\boldsymbol{w}_t'\boldsymbol{\Sigma}_{t+h|t}^{\mathrm{rx}}\boldsymbol{w}_t\right\}$$
$$\text{s.t. } \boldsymbol{w}_t'\boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}} = r_p^*$$

where $\boldsymbol{w}_t$ is a $(N, 1)$-vector of portfolio weights on risky zero-coupon bonds. Solving this optimization problem give us

$$\boldsymbol{w}_t = \frac{r_p^*}{C_t}(\boldsymbol{\Sigma}_{t+h|t}^{\mathrm{rx}})^{-1}\boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}}, \quad (16)$$

where $C_t = (\boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}})'(\boldsymbol{\Sigma}_{t+h|t}^{\mathrm{rx}})^{-1}\boldsymbol{\mu}_{t+h|t}^{\mathrm{rx}}$. Denote by $\boldsymbol{w}_t^{\mathcal{M}}$ the optimal weights at time t given by a model $\mathcal{M}$. The portfolio gross return given by $\mathcal{M}$ is

$$R_p^{\mathcal{M}} = 1 + r_f T_h + (\boldsymbol{w}_t^{\mathcal{M}})'\mathrm{rx}_{t+h}.$$

References

Bansal, R., & Zhou, H. (2002). Term structure of interest rates with regime shifts. *The Journal of Finance*, 57(5), 1997–2043.

Dai, Q., Singleton, K.J., & Yang, W. (2007). Regime shifts in a dynamic term structure model of US Treasury bond yields. *Review of Financial Studies*, 20(5), 1669–1706.

Della Corte, P., Sarno, L., & Thornton, D.L. (2008). The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. *Journal of Financial Economics*, 89(1), 158–174.

Elliott, R.J., & Siu, T.K. (2009). On Markov-modulated exponential-affine bond price formulae. *Applied Mathematical Finance*, 16(1), 1–15.

Feldhütter, P., & Nielsen, M.S. (2012). Systematic and idiosyncratic default risk in synthetic credit markets. *Journal of Financial Econometrics*, 10(2), 292–324.

Kikuchi, K., & Shintani, K. (2012). Comparative analysis of zero coupon yield curve estimation methods using JGB price data. *Monetary and Economic Studies*, 30, 75–121.

Kim, C.J., & Nelson, C.R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT press.

Koeda, J. (2013). Endogenous monetary policy shifts and the term structure: Evidence from Japanese government bond yields. *Journal of the Japanese and International Economies*, 29, 170–188.

Landen, C. (2000). Bond pricing in a hidden Markov model of the short rate. *Finance and Stochastics*, 4(4), 371–389.

Oda, N., & Ueda, K. (2007). The effects of the Bank of Japan's zero interest rate commitment and quantitative monetary easing on the yield curve: A macro-finance approach. *Japanese Economic Review*, 58(3), 303–328.

Roberts, G.O., Gelman, A., Gilks, W.R., et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1), 110–120.

Sim, D., & Ohnishi, M. (2014). A modified arbitrage-free Nelson–Siegel model: An alternative affine term structure model of interest rates. *Asia-Pacific Financial Markets*, (pp. 1–22).

Steeley, J.M. (1991). Estimating the gilt-edge term structure: basis splines and confidence intervals. *Journal of Business Finance & Accounting*, 18(4), 513–529.

Thornthon, D.L., & Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *The Review of Financial Studies*, 25(10), 3141–3168.

Wu, S., & Zeng, Y. (2007). An exact solution of the term structure of interest rate under regime-switching risk. In *Hidden Markov Models in Finance*, (pp. 1–14), Springer.

Xiang, J., & Zhu, X. (2013). A regime-switching Nelson–Siegel term structure model and interest rate forecasts. *Journal of Financial Econometrics*, 11(3), 522–555.

Zantedeschi, D., Damien, P., & Polson, N.G. (2011). Predictive macro-finance with dynamic partition models. *Journal of the American Statistical Association*, 106(494), 427–439.

Graduate School of Economics
Osaka University, Toyonaka 560-0043, Japan
E-mail address: simdara@gmail.com

大阪大学・経済学研究科 SIM DARA

Graduate School of Economics
Osaka University, Toyonaka 560-0043, Japan
E-mail address: ohnishi@econ.osaka-u.ac.jp

大阪大学・経済学研究科 大西 匡光