#2006-15

Discussion Paper #2006-15

Time Series Analysis of the Expectations Hypothesis for the Japanese Term Structure of Interest Rates in the Presence of Multiple Structural Breaks

by

Katsuhiro Sugita

Time Series Analysis of the Expectations Hypothesis for the Japanese Term Structure of Interest Rates in the Presence of Multiple Structural Breaks

Katsuhiro Sugita *

Abstract

This paper investigates the expectations hypothesis for the Japanese term structure of interest rates using vector error correction models with multiple structural breaks, focusing on how the breaks affect volatility, risk premium and speed of the adjustment toward the equilibrium. Using 1985-2005 data, we find strong evidence of three structural changes. After the second break point, the term structure relationship is found to be weakened with nearly zero percent short-term interest rate. This finding is consistent with the expectations hypothesis since with very low short-term interest rate the risk premium is dominant in determining long rates.

Key words: Term structure; Structural break; Cointegration; Bayesian inference; Gibbs sampling; Bayes factor;
JEL classification: C11; C13; C32; E43;

* Address: Graduate School of Economics, Hitotsubashi University, 2-1, Naka, Kunitachi, Tokyo 186-8601, Japan, Email: ksugita@econ.hit-u.ac.jp, Tel&Fax: (+81)(0)42 580-8525

1 Introduction

In the past few decades the Japanese economy has swung from rapid expansion in the late 80's to recession after the 1990 burst of the economic bubble. Japanese interest rates have moved in tandem. Interest rates exhibited a declining trend up to 1987 when both long- and short-term rates fell to around 4 percent, causing rapid economic expansion. The Bank of Japan (BoJ) introduced higher interest rates policy to alleviate the economic expansion, and maintained its higher interest rates policy until 1991. In 1990 the economic bubble burst, and then interest rates dropped rapidly with the BoJ's implementation of expansionary policies in an attempt to stimulate the sluggish economy. In 1999 the BoJ introduced the *zero-interest-rate-policy* that caused the short-term interest rates to approach almost zero percent yield. Thus, the movement of the Japanese interest rates are largely affected by the BoJ's monetary policy, and considering a time series model without multiple structural breaks might be misleading.

This paper investigates the expectations hypothesis for the Japanese term structure of interest rates using vector error correction models with multiple structural breaks in deterministic terms, adjustment terms, risk premium and covariance-variance matrices. The term structure of interest rates implies a stable relationship between interest rates with different maturities. This stable relationship would not be maintained when the short rate is successively close to zero percent because with lower short-term rate the spread between the two different rates approaches the risk premium and contains less information to account for the future long-term rates (Nagayasu, 2004). We provide a simple methodology for empirical evidence of the expectations hypothesis using cointegrated VAR models with multiple structural breaks.

The cointegration analysis for U.S. term structure has received significant attention since Campbell and Shiller (1987). Hall *et al*. (1992), Engsted and Tanggaard (1994), and Pagan *et al*. (1996) did not reject the null of no cointegration. However, after accounting for structural break(s), Bliss and Smith (1998), Lanne (1999), Hansen and Johansen (1999), Hansen (2003) detected the cointegration relationship. Aside from structural break, Sola and Driffill (1994) detected the cointegration of the expectations hypothesis using a Markov switching model. Other nonlinear examples are found in Dwyer *et al.* (1996), Tsay (1998), van Dijk and Franses (2000), Hansen and Seo (2001), Clements and Galvao (2001), and Clements and Galvao (2002).

To test for and estimate structural break in a multivariate dynamic model, Gregory and Hansen (1996a) developed residual-based tests for a single structural break, applying $ADF-$, $Z_\alpha-$, and Z_t- type tests to test the null of no cointegration against the alternative of cointegration in the presence of a possible regime shift. Gregory and Hansen (1996b) extended this to allow a trend shift in a cointegration model with a single break point. Hansen and Johansen (1999) tested a single structural break based on Nyblom's L statistic (1989). Bai *et al.* (1998) applied the Andrews-Ploberger exponential Wald statistic (1994) and developed methods for constructing confidence intervals for the date of a single break in multivariate time series. Seo (1998) provided the Lagrange multiplier test for structural breaks in a cointegration model, applying $supLM$ statistic by Andrews and Ploberger (1994). Testing for structural breaks in any subset of parameters in cointegration models was proposed by Hansen (2003), but this method assumed that the location of the break points are known.

This paper applies a Bayesian approach to analyze a vector error correction model for the Japanese term structure model, extending Wang and Zivot's (2000) method for detecting multiple breaks in univariate models as a problem of model selection. The Bayesian method with Markov chain Monte Carlo simulation technique makes testing for and estimating of multiple structural breaks in cointegration models technically simpler. Furthermore, the Bayesian method provides useful posterior information such as posterior density and uncertainty in the location of the break points rather than just point estimation. While Bai *et al.* (1998) proposed a method for constructing confidence intervals for the date of a single break in a classical framework, the Bayesian method provides HPDIs (highest posterior density interval) for the dates of multiple breaks.

This paper is organized as follows. Section 2 discusses the expectations hypothesis for the term structure of the interest rate, and reviews briefly movement of the Japanese interest rates. Modeling and its estimation method are presented in Section 3. Results of the empirical estimation for the Japanese term structure are reported in Section 4. Section 5 concludes. All computation in this article was performed using Ox v3.40 for Linux (Doornik, 2001).

2 Expectations Hypothesis for the Term Structure of the Interest Rates

2.1 The Expectations Hypothesis

The expectations hypothesis for the term structure of the interest rates states that the f-period interest rate is equal to the weighted average of the expected one period return plus a risk premium. For an overview of the expectations hypothesis theory, see Shiller (1990). Let $r_{f,t}$ be the yield to maturity for an f-period at time t, $L_{f,t}$ be the risk premium for an f-period at time t, then the hypothesis implies:

$$r_{f,t} = f^{-1}\sum_{i=1}^{f} E_t r_{1,t+i-1} + L_{f,t} \tag{1}$$

By rewriting the above equation, the interest rate spread $S_{f,t}$ can be expressed as:

$$S_{f,t} \equiv r_{f,t} - r_{1,t} = f^{-1}\sum_{i=1}^{f-1}\sum_{j=1}^{i} E_t \Delta r_{1,t+j} + L_{f,t} \tag{2}$$

If $r_{1,t}$ is integrated of order one, then $r_{f,t}$ is also integrated of order one and thus $r_{f,t}$ and $r_{1,t}$ are cointegrated with cointegrating vector (1, -1) as analyzed by Campbell and Shiller (1987). The risk premium is assumed to be $I(0)$ so that the hypothesis states that $r_{f,t} - r_{1,t} - L_{f,t}$ is a stationary process. Equation (2) states that if the change in expected short-term rate is zero percent, the spread $S_{f,t}$ equals the risk premium $L_{f,t}$.

The expectations hypothesis in equation (2) with constant risk premium implies the following vector error correction model with p lags:

$$\Delta r_t = D_t + \alpha(r_{f,t-1} - r_{1,t-1} - L_{f,t}) + \sum_{i=1}^{p-1} \Psi_i \Delta r_{t-i} + \varepsilon_t \tag{3}$$

where $r_t = \begin{pmatrix} r_{f,t} & r_{1,t} \end{pmatrix}'$, D_t is the deterministic term, α (2×1) is the speed of the adjustment term, Ψ_i (2×2) is the lag coefficient, and ε_t (2×2) is $iid\mathrm{N}\begin{pmatrix} 0 & \Omega \end{pmatrix}$. In this paper the risk premium $L_{f,t}$ is assumed to be either constant such as $L_t = \delta$ (as in Hansen, 2003) or constant with trend term such as $L_t = \delta + \gamma t$ within a given regime.

2.2 Japanese Interest Rates and the Expectations Hypothesis

Figure 1 illustrates the movement of the Japanese short-term interest rate (3-month bill rate) and the long-term interest rate (5-year government bond yield) since 1985. Until 1987, both short- and long-rates were decreasing partly because the Japanese government used monetary policy to ensure an appreciation of the Japanese Yen to avoid more accumulation of the trade surpluses against the US. The lower interest rates boosted the Japanese economy hereafter. The Bank of Japan (BoJ) introduced a higher interest rate policy to catch up with rapid economic expansion. It was, however, too late to prevent the economy from overheating, and resulted in bubble economy. The bubble economy burst in 1990, and accordingly, the interest rates exhibited a declining trend until the BoJ introduced the *zero-interest-rate policy* (ZIRP) between 1999 and 2001 in order to provide adequate liquidity. Under the ZIRP, the BOJ maintained a nearly zero percent overnight call rate. In 2001, the BoJ implemented a different operating target, which is the so called *quantitative-easing policy* (QEP). The aim was at further expansionary monetary policy by injecting liquidity into the market by setting the level of its current account as the operating target instead of targeting the level of the overnight call rate. Figure 1 shows that the short rate has been closed to zero since the ZIRP was implemented in 1999 and even since the QEP was implemented in 2001. The expectation hypothesis in equation (2) shows that with 0 percent of the expected short term rates the size of the yield spread equals the risk premium, which is assumed $I(0)$, that is, the long-term rate is merely the risk premium in this period, and thus the cointegration relationship between the two rates does not occur in this period. Since the ZIRP was introduced by the BoJ, the short-term interest rate has been kept nearly zero percent so that the cointegrating relationship between the two rates might have been vanished.

Figure 2 plots the yield spread between Japanese short- and long-term interest rates. If the expectations hypothesis holds, the spread shown in Figure 2 should follow a stationary process. The spread shows a negative trend until around 1991. This negative trend implies a lower risk premium due to higher expectation of future economic expansion. After 1991 it moved in an upward trend with higher risk premium until around 1996. In 1996 the short-rate suddenly dropped and the long-rate decreased gradually, so the spread decreased until 1999. After the 1999 implementation of the ZIRP, the spread has seemed to be stable due to the stationary process of the long-term interest rate that equals the risk premium ($I(0)$) according to the expectation hypothesis.

Table 1 presents the results of several cointegration tests: PIC (Chao and Phillips, 1999), Kleibergen & Paap's Bayesian test (KP) with diffuse prior (Kleibergen and Paap, 2002), Strachan and Inder's Grassman approach (2004), and Johansen's LR test (1991).[1] The PIC selects the rank 0 that has the smallest statistic. The KP test shows that the posterior probability of rank 0 is the highest with 85.1%. The Strachan's method also chooses rank 0 with 97.9%. The Johansen's trace test cannot reject the null of $r = 0$ at 5 percent significant level. Thus, all five tests cannot detect cointegration relationship between Japanese long and short term interest rates if structural breaks are not included in the model.

3 A Time Series Model with Multiple Structural Breaks in a Cointegrated VAR Model

3.1 Statistical Model

To investigate how structural breaks affect the speed of adjustment, the cointegrating vector, the risk premium and other terms, we consider the form of a vector error correction model with multiple structural breaks in equation (3). It is possible to consider a more general model where lag terms also change with breaks; however, for parsimonious reason, we assume that the lag terms do not change over time.

Let X_t denote an $I(1)$ vector of n-dimensional time series. The long-run multiplier matrix is decomposed as $\alpha\beta'$, both are $n \times r$, where α is the adjustment term and β' is the cointegrating vector. In this paper $X_t = (r_{l,t}, r_{s,t})$ where $r_{l,t}$ denotes long-term interest rate and $r_{s,t}$ short-term

[1] The PIC and the KP methods require r^2 (r is the number of rank in the long-run multiplier matrix) linear restrictions on the cointegrating vector for identification and normalization. Strachan (2003) and Strachan and Inder (2004) criticized this linear normalization as likely invalid. See Koop et al (2004) for a general survey of Bayesian cointegration analysis with a focus on the prior elicitation.

interest rate, and the cointegrating vector β is defined as $(1,-1,-\delta_t,-\gamma_t)$ so that the long-run equilibrium is represented as $S_{t-1}-(\delta_t+\gamma_t t)$ where $S_{t-1}\equiv r_{l,t-1}-r_{s,t-1}$ is the spread of the two interest rates, $r_{l,t}$ and $r_{s,t}$, at $t-1$ period, $\delta_t+\gamma_t t$ denotes the risk premium under the rational expectations hypothesis. If we assume that all parameters except the lag terms in VECM are subject to structural breaks, then the bivariate VECM representation is:

$$\Delta X_t=\mu_t+\xi_t t+\alpha_t\left(S_{t-1}-\delta_t-\gamma_t t\right)+\sum_{i=1}^{p-1}\Psi_i\Delta X_{t-i}+\varepsilon_t \tag{4}$$

where $t=p,p+1,\ldots,T$, and p is the number of lags, and ε_t are assumed $N(0,\Omega_t)$ and independent over time. Dimensions of matrices are μ_t, ξ_t and ε (2×1), $\Psi_{t,i}$ and Ω_t (2×2). We assume that the parameters μ_t, ξ_t and Ω_t are subject to $m<t$ structural breaks with break points $k_1,\ldots,k_m$, where $k_1<k_2<\cdots<k_m$, so that the observations can be separated into $m+1$ regimes.

Equation (4) can be rewritten in the matrix format as:

$$Y=WB+E \tag{5}$$

where

$Y=\begin{bmatrix}\Delta r'_p & \Delta r'_{p+1} & \cdots & \Delta r'_T\end{bmatrix}'$, $E=\begin{bmatrix}\varepsilon'_p & \varepsilon'_{p+1} & \cdots & \varepsilon'_T\end{bmatrix}'$

$W=\begin{bmatrix}X & Z_1 & \cdots & Z_{m+1}\end{bmatrix}$, $B=\begin{bmatrix}\Gamma' & \alpha'_1 & \cdots & \alpha'_{m+1}\end{bmatrix}'$,

$$Z_i=\begin{bmatrix} s_{i,p-1}(S_{p-1}-\delta_p-\gamma_p p)\\ s_{i,p}(S_p-\delta_{p+1}-\gamma_{p+1}(p+1))\\ \vdots \\ s_{i,T-1}(S_{T-1}-\delta_T-\gamma_T T)\end{bmatrix} \text{ for } i=1,\ldots,m+1,$$

$\Gamma=\begin{bmatrix}\mu_1 & \cdots & \mu_{m+1} & \xi_1 & \cdots & \xi_{m+1} & \Psi_1 & \cdots & \Psi_{p-1}\end{bmatrix}$,

$$X=\begin{bmatrix} s_{1,p} & \cdots & s_{m+1,p} & s_{1,p} & \cdots & s_{m+1,p} & \Delta r'_{p-1} & \cdots & \Delta r'_1 \\ s_{1,p+1} & \cdots & s_{m+1,p+1} & 2s_{1,p+1} & \cdots & 2s_{m+1,p+1} & \Delta r'_p & \cdots & \Delta r'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{1,T} & \cdots & s_{m+1,T} & (T-p+1)s_{1,T} & \cdots & (T-p+1)s_{m+1,T} & \Delta r'_{T-1} & \cdots & \Delta r'_{T-p+1} \end{bmatrix}.$$

Let τ be the number of rows of Y, so that $\tau = T-p+1$, then X is $\tau \times 2(m+p)$, Γ is $2(m+p)\times 2$, W is $\tau \times \kappa$ where $\kappa = 3m+2p+1$, and B is $\kappa \times 2$. $s_{i,j}$ in X is an indicator variable that equals 1 if the regime is i and 0 otherwise. Equation (5) represents the multivariate regression format of equation (4).

3.2 Prior Distributions and Likelihood Functions

We specify the proper prior distributions for the parameters given in the model (4). Let $k = (k_1, k_2, \ldots, k_m)'$ denote the vector of break dates. For the prior k, we choose a prior that is uniform over all ordered subsequences of $t = p+1, \ldots, T-1$. For priors for the risk premium terms, let $\eta_i = (\delta_i, \gamma_i)'$ and σ_i^2, $i = 1, \ldots, m+1$, be the error variance in the linear regression of the long run equilibrium $S_t = \delta_t + \gamma_t t + u_t$, $u_t \sim iidN(0, \sigma^2)$, then the prior for these parameters are such that the joint prior $p(\eta_i \mid \sigma_i^2)p(\sigma_i^2) = p(\eta_i, \sigma_i^2)$ is the normal inverted gamma density. For the prior for B, we consider that the vectorized B is the normal unconditional on Ω_i. We assume prior independence between k, B, Ω_i, and $\left(\eta_i, \sigma_i^2\right)$, $i = 1, 2, \ldots, m$+1, such that

$$p\left(k, B,\ \Omega_1, \ldots, \Omega_{m+1}, \eta_1, \ldots, \eta_{m+1}, \sigma_1^2, \ldots, \sigma_{m+1}^2\right) = p\left(k\right) p\left(B\right) \textstyle\prod_{i=1}^{m+1} \left\{p\left(\Omega_i\right) p\left(\eta_i, \sigma_i^2\right)\right\}.$$

The priors for k, Ω_i, $vec(B)$, and η_i are given as follows:

$$k \sim uniform(p+1, T-1) \tag{6}$$

$$\Omega_i \sim IW\left(\Lambda_i, h_i\right) \tag{7}$$

$$vec(B) \sim N(vec(B_0), V_0) \tag{8}$$

$$(\eta_i, \sigma_i^2) \sim NIG(\eta_{0,i}, M_{0,i}, s_{0,i}, \nu_{0,i}) \tag{9}$$

where IW refers to an inverted Wishart distribution with parameters $\Lambda_i \in \mathbb{R}^{2\times 2}$ and degrees of freedom, h_i; N refers to a multivariate normal with mean $vec(B_0) \in \mathbb{R}^{2\kappa\times 1}$ and covariance $V_0 \in \mathbb{R}^{2\kappa\times 2\kappa}$ in (8); NIG denotes a normal-inverted gamma density with mean $\eta_{0,i} \in \mathbb{R}^{2\times 1}$, covariance $M_{0,i} \in \mathbb{R}^{2\times 2}$, σ_i^2, $s_{0,i}$ and $\nu_{0,i}$ are scalar in (9).

Parameters for the risk premium, $\eta_i = (\delta_i, \gamma_i)'$, are assumed to be independent from parameters such as B and Ω_i in the VECM but dependent upon the break point k_{i-1} and k_i. Thus, η_i is derived from a simple regression $S_{i,t} = \delta_i + \gamma_i t + e_t = z_t \eta_i + e_{i,t}$ conditional on k_{i-1} and k_i where $S_{i,t}$ is the subsample of the regime i, $z_t = (1,t)$, and $e_{i,t}$ is the Gaussian error term such as $e_{i.t} \sim iid(0, \sigma_i^2)$ under the condition of the stationarity from the expectations hypothesis. It is, therefore, considered as a conventional Bayesian linear regression model such that if the natural conjugate prior with normal-inverted gamma density is assigned, then the marginal posterior density of η_i is a Student-t distribution, thus the posterior can be obtained analytically.

The joint prior of k, B, Ω_i, η_i and σ_i^2 is given by multiplication of (6) - (9) as follows:

$$\begin{aligned}
& p\left(k, B, \Omega_1, \ldots, \Omega_{m+1}, \eta_1, \ldots, \eta_{m+1}, \sigma_1^2, \ldots \sigma_{m+1}^2\right) \\
& = p(k, B, \Omega_1, \ldots, \Omega_{m+1})\, p\left(\eta_1, \ldots, \eta_{m+1}, \sigma_1^2, \ldots \sigma_{m+1}^2\right) \\
& \propto \left(\prod_{i=1}^{m+1} |\Lambda_i|^{h_i/2} |\Omega_i|^{-(h_i+n+1)/2}\right) |V_0|^{-1/2} \\
& \times \exp\left[-\frac{1}{2}\left\{\mathrm{tr}\left[\sum_{i=1}^{m+1}\left(\Omega_i^{-1}\Lambda_i\right)\right] + vec(B-B_0)' V_0^{-1} vec(B-B_0)\right\}\right] \\
& \times \exp\left(-\frac{1}{2}\sum_{i=1}^{m+1}\left[\sigma_i^{-2}\left\{s_{0,i} + (\eta_i - \eta_{0,i})' M_{0,i} (\eta_i - \eta_{0,i})\right\}\right]\right).
\end{aligned} \tag{10}$$

The likelihood function for k, B, Ω_i, η_i and σ_i^2 is given by,

$$\begin{aligned}
&\mathfrak{L}\left(k, B, \Omega_1, \ldots, \Omega_{m+1}, \eta_1, \ldots, \eta_{m+1}, \sigma_1^2, \ldots \sigma_{m+1}^2 \mid Y\right) \\
&\propto \left(\prod_{i=1}^{m+1} |\Omega_i|^{-t_i/2} \sigma_i^{-t_i}\right) \exp\left(-\frac{1}{2}\mathrm{tr}\left[\sum_{i=1}^{m+1}\left\{\Omega_i^{-1}(Y_i - W_iB)'(Y_i - W_iB)\right\}\right]\right) \\
&\times \exp\left(-\frac{1}{2}\sum_{i=1}^{m+1}\left[\sigma_i^{-2}(S_i - Z_i\eta_i)'(S_i - Z_i\eta_i)\right]\right) \\
&= \left(\prod_{i=1}^{m+1} |\Omega_i|^{-t_i/2} \sigma_i^{-t_i}\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{m+1}\left[(vec(Y_i - W_iB))'(\Omega_i \otimes I_\tau)^{-1}(vec(Y_i - W_iB))\right]\right) \\
&\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{m+1}\left(\sigma_i^{-2}\left[\zeta_i + (\eta_i - \widehat{\eta}_i)'Z_i'Z_i(\eta_i - \widehat{\eta}_i)\right]\right)\right\} \qquad (11)
\end{aligned}$$

where Y_i denotes the $t_i \times 2$ submatrix of Y values in regime i, W_i denotes $t_i \times \kappa$ submatrix of W in regime i, and t_i is the number of observations in regime i when $s_t = i$, $i = 1, 2, \ldots, m+1$, $S_i = \{S_{i,1}, S_{i,2}, \ldots, S_{i,t_i}\}'$, $\widehat{\eta}_i = (Z_i'Z_i)^{-1}Z_i'S_i$, $\zeta_i = S_i'(I_{t_i} - Z_i(Z_i'Z_i)^{-1}Z_i)S_i$, $Z_i = (z_1, z_2, \ldots, z_{t_i})'$.

3.3 Posterior Specifications and Estimation

The joint posterior distribution can be obtained from the joint priors given in (10) multiplied by the likelihood function for k, B, and Ω_i that is,

$$\begin{aligned}
&p\left(k, B, \Omega_1, \ldots, \Omega_{m+1}, \eta_1, \ldots, \eta_{m+1}, \sigma_1, \ldots \sigma_{m+1} \mid Y\right) \\
&= p\left(k, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y\right) p\left(\eta_1, \ldots, \eta_{m+1}, \sigma_1, \ldots \sigma_{m+1} \mid k, Y\right) \\
&\propto p\left(k, B, \Omega_1, \ldots, \Omega_{m+1}\right) \mathfrak{L}\left(k, B, \Omega_1, \ldots, \Omega_{m+1} \mid Y\right) \\
&\times p\left(\eta_1, \ldots, \eta_{m+1}, \sigma_1, \ldots \sigma_{m+1}\right) \mathfrak{L}\left(\eta_1, \ldots, \eta_{m+1}, \sigma_1, \ldots \sigma_{m+1} \mid Y\right) \\
&\propto \left(\prod_{i=1}^{m+1}\left\{|\Lambda_i|^{h_i/2} |\Omega_i|^{-(t_i+h_i+n+1)/2} \sigma_i^{-t_i}\right\}\right) |V_0|^{-1/2} \\
&\times \exp\left(-\frac{1}{2}\left[\mathrm{tr}\left(\sum_{i=1}^{m+1}\Omega_i^{-1}\Lambda_i\right) + \sum_{i=1}^{m+1}\left\{\left([vec(Y_i - W_iB)]'(\Omega_i \otimes I_\tau)^{-1} vec(Y_i - W_iB)\right)\right\}\right.\right.
\end{aligned}$$

$$+vec\,(B-B_0)'V_0^{-1}vec\,(B-B_0)\big]\Big)$$
$$\times\exp\left\{-\frac{1}{2}\sum_{i=1}^{m+1}\left[\sigma_i^{-2}\left\{s_{0,1}+\zeta_i+(\eta_i-\eta_{0,i})'M_{0,i}\,(\eta_i-\eta_{0,i})+(\eta_i-\widehat{\eta}_i)'Z'Z\,(\eta_i-\widehat{\eta}_i)\right\}\right]\right\}. \tag{12}$$

Consider first the conditional posterior of k_i, $i=1,2,\dots,m$. Given that $1=k_0<\cdots<k_{i-1}<k_i<k_{i+1}<\cdots<k_{m+1}=\tau$ and the form of the joint prior, the sample space of the conditional posterior of k_i only depends on the neighboring break dates k_{i-1} and k_{i+1}. It follows that, for $k_i\in[k_{i-1},k_{i+1}]$,

$$p\,(k_i\mid[\Theta-k_i],Y)\propto p\,(k_i\mid k_{i-1},k_{i+1},B,\Omega_i,\Omega_{i+1},\eta_i,\eta_{i+1},Y) \tag{13}$$

for $i=1,\dots m$, which is proportional to the likelihood function for $\Theta=(k',B',\Omega_1',\dots\Omega_{m+1}',\eta_1',\dots,\eta_{m+1}')'$ evaluated with a break at k_i only using data between k_{i-1} and k_{i+1} and probabilities proportional to the likelihood function.

Next, we consider the conditional posterior of Ω_i and $vec\,(B)$. From the joint posterior in (12), we can write two terms as:

$$\sum_{i=1}^{m+1}\left\{[vec\,(Y_i-W_iB)]'\,(\Omega_i\otimes I_\tau)^{-1}\,vec\,(Y_i-W_iB)\right\}+[vec\,(B-B_0)]'V_0^{-1}vec\,(B-B_0)$$
$$=[vec\,(B-B_\star)]'V_B^{-1}vec\,(B-B_\star)+Q$$

where

$$Q=\sum_{i=1}^{m+1}\left\{[vec\,(Y_i)]'\,(\Omega_i\otimes I_\tau)^{-1}\,vec\,(Y_i)\right\}+[vec\,(B_0)]'V_0^{-1}vec\,(B_0)-[vec\,(B_\star)]'M_\star^{-1}vec\,(B_\star).$$

Thus, the conditional posterior of Ω_i is derived as an inverted Wishart distribution as:

$$p(\Omega_i\mid k,B,\eta_i,Y)\propto|\Xi_{i,\star}|^{t_i/2}|\Omega_i|^{-(t_i+h_i+n+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega_i^{-1}\Xi_{i,\star}\right)\right] \tag{14}$$

where $\Xi_{i,\star}=(Y_i-W_iB)'\,(Y_i-W_iB)+\Lambda_i$. The conditional posterior of $vec\,(B)$ is derived as a multivariate normal density with covariance, V_B, that is,

$$\begin{aligned}
&p(vec\,(B) \mid k, \eta_1, \ldots, \eta_{m+1}, \Omega_1, \ldots, \Omega_{m+1}, Y) \\
&\propto |V_B|^{-1/2} \exp\left[-\frac{1}{2}\left\{[vec\,(B - B_\star)]' V_B^{-1} vec\,(B - B_\star)\right\}\right]
\end{aligned} \tag{15}$$

where

$$vec\,(B_\star) = \left[V_0^{-1} + \sum_{i=1}^{m+1}\left\{\Omega_i^{-1} \otimes \left(W_i' W_i\right)\right\}\right]^{-1}\left[V_0^{-1} vec\,(B_0) + \sum_{i=1}^{m+1}\left\{(\Omega_i \otimes I_k)^{-1} vec\left(W_i' Y_i\right)\right\}\right].$$

and

$$V_B = \left[V_0^{-1} + \sum_{i=1}^{m+1}\left\{\Omega_i^{-1} \otimes \left(W_i' W_i\right)\right\}\right]^{-1}.$$

The posterior of η_i is a Student-t density conditional on k that is derived analytically from the joint posterior with σ_i^2 (a normal-inverted gamma density) as the following:

$$\eta_i \sim t(\overline{\eta}_i, s_{*i}, M_{*i}, \nu_{*i}) \tag{16}$$

where $M_{*i} = M_{0i} + \Upsilon_i' \Upsilon_i$, $\overline{\eta}_i = M_{*i}^{-1}(M_{0i}\eta_{0i} + \Upsilon_i' \Upsilon_i \widehat{\eta}_i)$, $s_{*i} = s_{0i} + \zeta_i + (\eta_{0i} - \widehat{\eta}_i)'[M_{0i}^{-1} + (\Upsilon_i' \Upsilon_i)^{-1}]^{-1}(\eta_{0i} - \widehat{\eta}_i)$, $\nu_{*i} = \nu_{0i} + t_i$. Thus, the posterior mean of η_i can be obtained as $E[\eta_i \mid y] = \overline{\eta}_i$.

Given the full set of conditional posterior specifications above, we illustrate the Gibbs sampling algorithm for generating sample draws from the posterior. The following steps can be replicated:

- Step 1: Set $j = 1$. Specify starting values for the parameters of the model, $k^{(0)}$, $B^{(0)}$, and $\Omega_i^{(0)}$ for $i = 1, 2, \ldots, m+1$.
- Step 2a: Compute likelihood probabilities sequentially for each date at $k_1 = k_0^{(j-1)} + 1, \ldots, k_2^{(j-1)} - 1$ to construct a multinomial distribution. Weight these probabilities such that the sum of them equals 1.

- Step 2b: Generate a draw for the first break date k_1 as a multinomial random variable on the sample space $\left[k_0^{(j-1)}, k_2^{(j-1)}\right]$ from

$$p\left(k_1^{(j)} \mid k_0^{(j-1)}, k_2^{(j-1)}, B^{(j-1)}, \Omega_1^{(j-1)}, \Omega_2^{(j-1)}, \eta_1^{(j-1)}, \eta_2^{(j-1)}, Y\right).$$

- Step 3a: For $i = 3, \ldots, m+1$, compute likelihood probabilities sequentially for each date at $k_{i-1} = k_{i-1}^{(j-1)} + 1, \ldots, k_i^{(j-1)}$ to construct a multinomial distribution. Weight these probabilities such that their sum equals 1.

- Step 3b: Generate a draw of the $(i-1)$th break date $k_{i-1}^{(j)}$ from the conditional posterior $p\left(k_{i-1}^{(j)} \mid k_{i-2}^{(j-1)}, k_i^{(j-1)}, B^{(j-1)}, \Omega_{i-1}^{(j-1)}, \Omega_i^{(j-1)}, \eta_{i-1}^{(j-1)}, \eta_i^{(j-1)}, Y\right)$. Return to repeat Step 3a, but imposing the previously generated break date, to generate the next break date. Iterate until all breaks are generated.

- Step 4: Compute $\overline{\eta}_i^{(j)}$ as $E[\eta_i \mid y] = \overline{\eta}_i$ where $\overline{\eta_i}$ is calculated from the posterior in (16).

- Step 5: Generate $B^{(j)}$ from $p(vec(B) \mid k^{(j-1)}, \eta_1^{(j)}, \ldots, \eta_{m+1}^{(j)}, \Omega_i^{(j-1)}, \ldots, \Omega_{m+1}^{(j-1)}, Y)$ in (15).

- Step 6: Generate $\Omega_i^{(j)}$ from $p(\Omega_i \mid k^{(j-1)}, B^{(j)}, \eta_i^{(j)}, Y)$ for all $i = 1, \ldots, m+1$ in (14).

- Step 7: Set $j = j+1$, and go to Step 2.

Step 2 through to Step 7 are iterated N times to obtain the posterior densities. Note that the first L iterations are discarded in order to remove the effect of the initial values.

3.4 Determining the Number of Structural Breaks and Model Selection by Bayes Factors

Determining the number of structural breaks in a vector error correction model can be treated as a problem of model selection. In a Bayesian context, the posterior probabilities for all models under consideration are used for model selection. The posterior probabilities can be obtained by Bayes factors, defined by the ratio of marginal likelihoods as $BF_{jk} = p(y \mid \mathrm{M}_j)/p(y \mid \mathrm{M}_k)$, and there are several methods to calculate the Bayes factors (see Kass and Raftery, 1995). In this paper we choose the Schwarz's Bayesian information criterion (BIC) to approximate the Bayes factors.

The Schwarz BIC can give a rough approximation of the Bayes factors, which is easy to use and does not require evaluation of the prior distribution as Kass and Raftery (1995) noted. Wang and Zivot (2000) employed the Schwarz BIC to calculate the Bayes factors for detecting the number of structural breaks in a univariate context. The Schwarz BIC for model j can be obtained as

$$\mathrm{BIC}_j = -2\ln\mathfrak{L}\left(\widehat{\theta}_j \mid Y; \mathrm{M}_j\right) + q_j \ln(t) \tag{17}$$

where $\mathfrak{L}\left(\widehat{\theta}_j \mid Y; \mathrm{M}_j\right)$ denotes the likelihood function for model j; q_j denotes the total number of estimated parameters in the model j and M_j denotes the model indicator for model j. The likelihood function $\mathfrak{L}\left(\widehat{\theta}_j \mid Y; \mathrm{M}_j\right)$ is evaluated at $\widehat{\theta}_j$, the posterior means of the parameters for model j based on the output of the Gibbs sampler.

The Bayes factor for model k against model j can be approximated by

$$BF_{jk} \approx \exp\left[-0.5\left(\mathrm{BIC}_j - \mathrm{BIC}_k\right)\right] \tag{18}$$

With the prior odds, defined as $\Pr(\mathrm{M}_j)/\Pr(\mathrm{M}_k)$, the posterior odds can be obtained by multiplying the Bayes factor by the prior odds as $\mathrm{PosteriorOdds}_{jk} = BF_{jk} \times \mathrm{PriorOdds}_{jk}$. We compute the posterior odds for all possible models and then obtain the posterior probability for each model by

$$\Pr\left(\mathrm{M}_j \mid Y\right) = \frac{\mathrm{PosteriorOdds}_{jk}}{\sum_{m=1}^{n} \mathrm{PosteriorOdds}_{mk}} \tag{19}$$

where n is the number of models under consideration.

By using the Schwarz BIC to approximate the logarithm of the Bayes factor, it is easy to determine the number of breaks and other model specification such as whether the volatility is subject to structural breaks as a problem of model selection. In our case, we compute the Schwarz

BIC as

$$\text{BIC}_j \quad = \quad \text{-}2\ln\mathfrak{L}(k,B,\Omega_1,\ldots,\Omega_{m+1},\eta_1,\ldots,\eta_{m+1}\,|\,Y;\text{M}_j)+q_j\ln(T) \tag{20}$$

We compute BIC_j using the posterior modes of k_j for $j=1,\ldots,m$ and the posterior means of the remaining parameters based on the output of the Gibbs sampler.

Alternative methods for calculating the Bayes factor include using the harmonic mean of the likelihood as the marginal likelihood (Newton and Raftery, 1994), or using the Gibbs output to calculate the marginal likelihood (Chib, 1995). Compared with these methods, the BIC approach gives merely a rough approximation although it is consistent in determining the number of structural breaks as shown by Yao (1988) and Liu et al (1997).

4 Estimation Results

In this section, we analyze the Japanese term structure of interest rates using the cointegration models with multiple structural breaks outlined in the previous section. The data used in this empirical study are 3-month bill rate as the short-term interest rate and 5-year government bond yield as the long-term interest rate based on the monthly data taken from IMF's *International Financial Statistics* and *Datastream* respectively ranged from 1985:01 to 2005:10 with 250 observations, and are plotted in Figure 1. Figure 2 presents the spread between the two rates.

We consider the VECM with $\Delta r_t = (\Delta r_{l,t}, \Delta r_{s,t})$, where $r_{s,t}$ denotes the short-term interest rate and $r_{l,t}$ denotes the long-term interest rate, and estimate eight models with structural breaks in different subset of the parameters with the number of breaks $m=0,1,\ldots,5$. The number of lags p is 3 selected by Schwarz BIC. The Gibbs sampling is performed with 10,000 draws and the first 1,000 discarded. The prior hyperparameters are chosen as $\Lambda_i = 0.1I_2$, $h_i = 2.001$ $\forall i$ in (7), $B_0 = 0_{\kappa\times n}$, $V_0 = 0.1I_\kappa$ in (8), $\eta_{0,i} = 0_{2\times 1}$, $M_{0,i}=0$, $s_{0,i}=0.1$, $\nu_{0,i}=0.01$ $\forall i$ in (9). These choices of the hyperparmeters are relatively noninformative.

In this empirical study, we are interested in how the breaks affect the the adjustment terms, risk premium and covariance-variance matrices so that models under consideration allow these parameters to change with breaks. The following eight models with different specifications were estimated[2]:

$$\text{Model 1: } \Delta r_t = \mu_t + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t^*$$

$$\text{Model 2: } \Delta r_t = \mu_t + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t$$

$$\text{Model 3: } \Delta r_t = \alpha_t (S_{t-1} - \delta_t) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t^*$$

$$\text{Model 4: } \Delta r_t = \alpha_t (S_{t-1} - \delta_t) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t$$

$$\text{Model 5: } \Delta r_t = \alpha_t (S_{t-1} - \delta) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t^*$$

$$\text{Model 6: } \Delta r_t = \alpha_t (S_{t-1} - \delta) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t$$

$$\text{Model 7: } \Delta r_t = \mu_t + \alpha (S_{t-1} - \delta) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t^*$$

$$\text{Model 8: } \Delta r_t = \mu_t + \alpha (S_{t-1} - \delta) + \sum_{i=1}^{2} \Psi_i \Delta r_{t-i} + \varepsilon_t$$

where $\varepsilon_t^* \sim iidN(0, \Omega_t)$ for Model 1, 3, 5 and 7, and $\varepsilon_t \sim iidN(0, \Omega)$ for other models. Model 1 and Model 2 assume that there is no cointegration relationship between the two variables. Model 1 allows μ and Ω to change with breaks, while Model 2 assumes constant volatility. The rest of models assume that there exist one cointegration relationship between the two interest rates. In Model 3 α, δ and Ω are subject to change with breaks. Model 4 restricts Ω to being constant over the entire sample. Model 5 assumes that α and Ω shift with breaks while in Model 6 Ω does not shift. The speed of the adjustment toward the equilibrium in both Model 5 and 6 is subject to

[2] We also considered models that contains the time trend in the cointegrating relationship. However, the Bayes factors for these models are insignificantly small compared with the models considered in this paper.

change while the risk premium in these models is not affected by the breaks. Model 7 assumes that intercept term μ and Ω shift with breaks while in Model 8 Ω does not change with breaks. In both Model 7 and 8 changes in risk premium does not affect the speed of the adjustment as Model 5 and 6, but the intercept terms are subject to change with the breaks. To compute the posterior probabilities for the models with a various number of the breaks, the Bayes factors approximated by (18) and (20) are calculated, and the results are reported in Table 2. From these results, a cointegration exists once the structural breaks are considered as $\Pr(\text{Model } 1 \mid Y) + \Pr(\text{Model } 2 \mid Y) = 0.000$. The most appropriate number of the break is $m = 3$ since the posterior probability when $m = 3$, $\Pr(m = 3 \mid Y) = \sum_{i=1}^{8} \Pr(m = 3 \mid \text{Model } i, Y) = 0.854$, which is dominant. Clearly, the no-structural break model ($m = 0$) is rejected by the data as $\Pr(m = 0 \mid Y) = 0.000$. A cointegration is not detected if the structural breaks were not considered as shown in Table 1, although once the breaks were taken into consideration a model with cointegration is strongly favored. The results reported in Table 2 show that the models where covariance matrices, Ω_t, change with breaks (Model 1, Model 3, and Model 5) are strongly supported against homoscedastic models. A model with the highest posterior probability is Model 3 with 97.2 percent ($\Pr(\text{Model } 3 \mid Y) = \sum_{i=0}^{5} \Pr(\text{Model } 3 \mid m = i, Y) = 0.972$). Other models exhibit ignorably low posterior probabilities. Hence Model 3 with $m = 3$ is dominant over other models with $\Pr(\text{Model } 3 \mid m = 3, Y) = 0.836$, and thus we focus on this model to investigate.

The estimated break points and the 95% HPDI (Highest Probability Density Intervals) of each break point for Model 3 with $m = 3$ are reported in Table 3 and plotted in Figure 3. The posterior mode of the three structural breaks are 1991:4, 1999:4, and 2001:7. The first estimated break date seems closely associated with the burst of the bubbled economy in 1990, and the second break seems associated with the implementation of *the zero-interest-rate-policy* in March 1999. The third break date seems to correspond with the introduction of *the quantitative easing policy* in March 2001.

The estimates of parameters for Model 3 excluding the coefficients of the two lag terms of the vector error correction model with three structural breaks are reported in Table 4. The results show that there are significant changes in volatility Ω. For example, Ω_1 in the first regime is the largest and then becomes smaller as both rates approach zero. This is not surprising since with the lower interest rates as in the ZIRP the volatility tends to be smaller. On the other hand, the higher interest rates tends to fluctuate much more than lower rates. In the third and fourth regimes, the standard deviations of the covariances between the error terms of the long- and short-term interest rate are very high; in other words, the covariances between the two rates are not significant. This suggests that the movement of the long-term interest rate is almost independent of those of the short-term interest rate in the third and fourth regimes where the short rate has been kept as low as zero percent, while in other regimes (regime 1 and 2) the covariances between the long- and short-term interest rates are significantly different from zero.

The estimated speed of adjustment toward the equilibrium, α_i, clearly differs between the four regimes. The speed of adjustment for both short- and long-term interest rates are decreased after the second break date, the implementation of the *zero-interest-rate policy* in 1999. In the first and second regimes before the second break point the speed of both long- and short-term rates are significant with small standard deviations; however, after the second break, the speed declines and becomes insignificant. After the second break point, the adjustment speed for both long- and short-term interest rate approaches almost zero with fairly large standard deviations. This implies that the cointegration relationship between the two interest rates is weakened after the second break point. This is consistent with the expectations hypothesis of the term structure, which implies that when the interest rates are lower the stable relationship of the interest spread is weakened and the spread is merely risk premium when the short-term interest rate reaches zero percent. Figure 4 plots the posterior density of each α (alpha-ij where i denotes regime 1,2,3,4 and j=1 for the long-term, j=2 for the short-term interest rate), and shows that both densities of α for both in the third and fourth regimes (alpha-31, alpha-32, alpha-41 and alpha-42) contain zero, which suggests that the adjustment toward the stable relationship does not occur in the regimes of the *zero-interest-rate-policy* and the *quantitative easing policy*. These analysis by the HPDI are sensible but informal in contrast to posterior odds. To confirm this no-cointegration in the third and the fourth regime in a formal way, we compute the Bayes factor using the Savage-Dickey density ratio method with the restrictions $\alpha_3 = \alpha_4 = 0$. The Savage-Dickey density ratio is used for comparing nested models where one model is restricted (M1) and the other is unrestricted (M2). The Bayes factor comparing these two models by Savage-Dickey density ratio is given by:

$$BF_{12} = \frac{p\left(\alpha_3 = \alpha_4 = 0 \mid Y, \text{Model 3}\right)}{p\left(\alpha_3 = \alpha_4 = 0 \mid \text{Model 3}\right)} \tag{21}$$

The denominator of the RHS in (21) is easily calculated since $vec(\alpha_i)$ is a part of $vec(B)$ which is Normal. The numerator of the RHS in (21) cannot be calculated directly since we have the conditional posterior for $vec(B)$ (that is $vec(\alpha)$) which is Normal, but not the marginal posterior density. The Gibbs output, however, can be used for estimation of the marginal posterior. Let N be the total number of the Gibbs iterations, N_0 be the number of draws that is discarded to remove the effect of the initial values. Then, averaging $p\left(\alpha_3 = \alpha_4 = 0 \mid k^{(n)}, \eta_3^{(n)}, \eta_4^{(n)}, \Omega_3^{(n)}, \Omega_4^{(n)}, Y_{3-4}, \text{Model 3}\right)$ across the draws $k^{(n)}, \eta_3^{(n)}, \eta_4^{(n)}, \Omega_3^{(n)}, \Omega_4^{(n)}$ will yield an estimate of $p\left(\alpha_3 = \alpha_4 = 0 \mid Y, \text{Model 3}\right)$. To be precise, let $\Theta^{(n)} = \left(k^{(n)}, \eta_3^{(n)}, \eta_4^{(n)}, \Omega_3^{(n)}, \Omega_4^{(n)}\right)$ be the n-th draw from the Gibbs sampler, then

$$\begin{aligned} & \frac{1}{N - N_0} \sum_{n=N_0+1}^{N} p\left(\alpha_3 = \alpha_4 = 0 \mid \Theta^{(n)}, Y_{3-4}, \text{Model 3}\right) \\ \rightarrow \quad & p\left(\alpha_3 = \alpha_4 = 0 \mid Y, \text{Model 3}\right) \end{aligned} \tag{22}$$

as N goes to infinity. We compute the Bayes factor using (21) and (22) to compare the restricted model with the unrestricted model. The Bayes factor for this results in 8.842, which suggests that the restricted model of the no-cointegration in regime 3 and 4 is supported with 89.84%.

Table 4 also shows that the changes in the values of the mean δ in the cointegrating relationship are significantly affected by the breaks. This parameter expresses the risk premium according to the expectations hypothesis. It is negative until the first break when the bubble burst in 1991. It then becomes positive as the future uncertainty increases in the recession. After the second break the risk premium again decreases with the expectation of recovery from the long recession. Then the risk premium is slightly increased after the third break.

5 Conclusion

This paper developed a Bayesian approach for a cointegrated VAR model with multiple structural breaks in order to analyze the expectations hypothesis for the Japanese term structure of the interest rates, extending Wang and Zivot's (2000) approach for univariate models. The Gibbs sampling method simplifies the estimation of this model. The number of structural break points are selected by the posterior probability based on the estimation of the models given the number of possible break dates. The Bayesian approach provides useful information such as uncertainty in the location of the dates by the posterior mass function for each estimated break points.

We found strong evidence of three structural breaks during 1985 - 2005. These three breaks seemed to be associated respectively with the burst of economic bubble in 1990, the implementation of the BoJ's *zero-interest rate policy* in late 1999, and *the quantitative easing policy* in 2001. The speed of the adjustment toward the equilibrium is found to be affected by the breaks. The adjustment terms approach almost zero after the second break date for both long- and short-term interest rates, which implies that there was no cointegration relation in the third regime when the short-term interest rate was kept at nearly zero percent; that is, it did not respond to the movement of the long-term interest rate. The Bayes factor calculated by the Savage-Dickey density ratio supports no cointegration during these periods. This finding is consistent with the expectations hypothesis of the term structure model that implies no cointegration when interest rates are low because the risk premium is dominant in the yield spread between the two interest rates. We also found that the volatility and the risk premium were affected by these three breaks.

References

[1] Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**: 1383-1414.

[2] Bai, J., Lumsdaine, R.L., and Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* **65**: 395-432.

[3] Bliss, R.R., and Smith, D.C., 1998. The elasticity of interest rate volatility: Chan, Kaolyi, Longstaff and Sanders Revisited. *Journal of Risk* **1**: 21-46.

[4] Campbell, J.Y., and Shiller, R.J., 1987. Cointegration and tests of present value models. *Journal of Political Economy* **95**: 1062-1088.

[5] Chao.J., and Phillps, P.C.B., 1999. Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* **91**: 227-271.

[6] Chib, S, 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313-1321.

[7] Clements, M.P., and Galvao, A.B., 2001. A comparison of tests of nonlinear cointegration with an application to the predictability of US interest rates using the term structure. Mimeo, University of Warwick.

[8] Clements, M.P., and Galvao, A.B., 2002. Testing the expectations theory of the term structure of interest rates in threshold models. *Macroeconomic Dynamics*: 1-19.

[9] Doornik, J.A., 2001. *OX An object-oriented matrix programming language*, Timberlake Consultants Press., London.

[10] Dwyer, G.P., Locke, P., and Yu, W., 1996. Index arbitrage and nonlinear dynamics between the SP500 futures and cash. *Review of Financial Studies* **9**: 301-332.

[11] Engsted, T., Tanggaard, C., 1994. Cointegration and the US term structure. *Journal of Banking and Finance*: 167-181.

[12] Gregory, A.W., and Hansen, B.E., 1996a. Residuals based tests for cointegration in models with regime shifts. *Journal of Econometrics* **70**: 99-126.

[13] Gregory, A.W., and Hansen, B.E., 1996b. Practitioners corner, testing for cointegration in models with regime and trend shifts. *Oxford Bulletin of Economics and Statistics* **58**: 555-560.

[14] Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasure bill yield. *The Review of Economic and Statistics* **74**: 116-126.

[15] Hansen, P.R., 2003. Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics* **114**: 261-295.

[16] Hansen, H., Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR-models. *Econometrics Journal* **2**: 306-333.

[17] Hansen, B.E., and Seo, B., 2001. Testing for two-regime threshold cointegration in vector error correction models. *Journal of Econometrics* **110: 293-318.**

[18] Johansen, S., 1991. The power function for the likelihood ratio test for cointegration. in J. Gruber, ed., *Econometric Decision Models: New Methods of Modelling and Applications.* Springer Verlag, New York, NY: 323-335.

[19] Kass, R.E., and Raftery, A. E., 1995. Bayes factors. *Journal of the American Statistical Association* **90**: 773-795.

[20] Kleibergen, F., and Paap, R. 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* **111**: 223-249.

[21] Koop, G., Strachan, R, van Dijk, H., Villani, M., 2004. Bayesian approaches to cointegration. Working Paper No.04/27, University of Leicester.

[22] Lanne, M., 1999. Near unit roots and the predictive power of yield spreads for changes in long-term interest rates. *The Review of Economic and Statistics* **81**: 393-398.

[23] Liu, J., Wu, S., and Zidek, J.V., 1997. On segmented multivariate regressions. *Statistica Sinica* **7**: 497-525.

[24] Nagayasu, J., 2004. Term structure of interest rates and monetary policy during a zero-interest-rate period. *Monetary and Economic Studies* **22**: 19-43.

[25] Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society* B **56**: 3-48.

[26] Nyblom, J., 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* **84**: 223-230.

[27] Pagan, A.R., Hall, A.D., Martin, V., 1996. Modelling the term structure. Handbook of Statistics **14**: 91-118.

[28] Seo, B., 1998. Tests for structural change in cointegrated system. *Econometric Theory* **14**: 222-259.

[29] Shiller, R.J., 1990. The term structure of interest rates. In: Friedman, B., Hahn, F. (eds), Handbook of Monetary Economics vol. 1., North-Holland, Amsterdam: 627-722.

[30] Sola, M, Driffill, J., 1994. Testing the term structure of interest rates using a stationary vector autoregression with regime switching. *Journal of Economics Dynamics and Control* **18**: 601-628.

[31] Strachan, R., 2003. Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics* **21**: 185-195.

[32] Strachan, R., and Inder, B., 2004. Bayesian analysis of the error correction model. *Journal of Econometrics* **123**: 307-325.

[33] Tsay, R.S., 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188-1202.

[34] van Dijk, D., Franses, P.H., 2000. Nonlinear error-correction models for interest Rates in the Netherlands. In W.A.Barnett, D.F.Hendry, S.Hylleberg, T.Terasvirta, D.Tjostheim, and A. Wurtz (eds.) *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory.* Cambridge, UK: Cambridge University Press.

[35] Wang, J., and Zivot, E., 2000. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business and Economic Statistics* **18** 3: 374-386.

[36] Yao, Y-C., 1988. Estimating the number of changepoints via Schwarz criterion. *Statistics and Probability Letters* **6**: 181-189.

Figure 1: Japanese long-term and short-term interest rates

solid line - long-term interest rate, dotted line - short-term interest rate

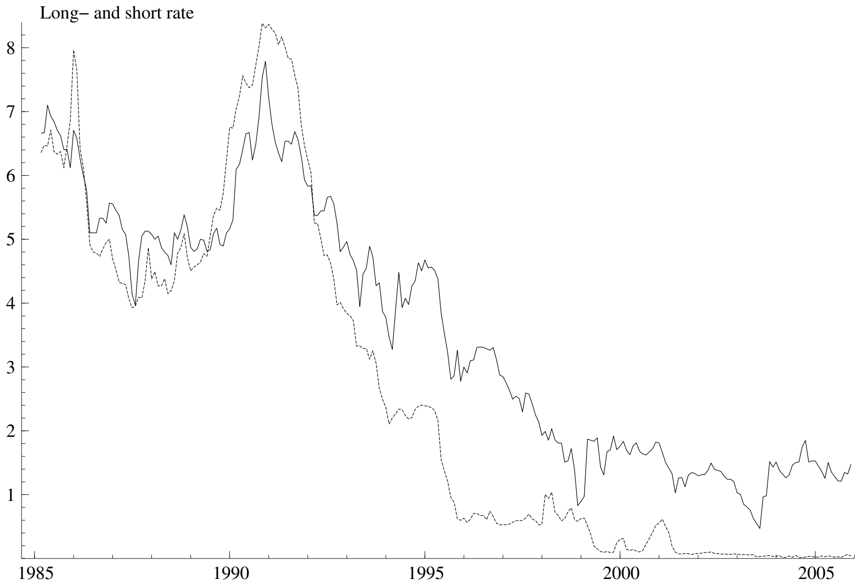

Figure 2: Spread between Long- and Short-term Interest Rates

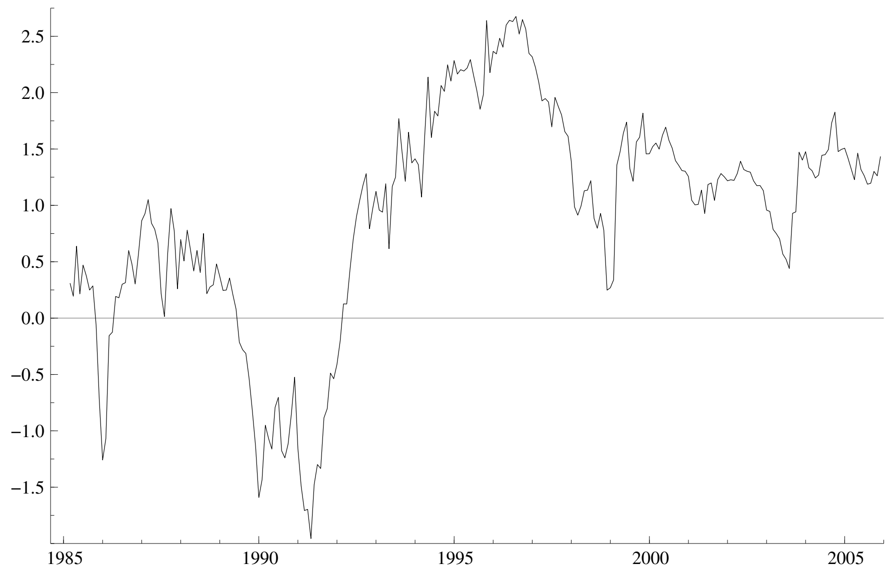

Table 1: Cointegration Tests

rank r	PIC	KP (%)	Strachan (%)	LR	5% cv
0	-2.174	0.851	0.979	8.209	12.53
1	-0.588	0.122	0.021	2.028	3.84
2	-0.127	0.027	0.000	–	–

Note:
The lag was chosen to be $p = 3$ in VAR by the BIC.
PIC: Posterior Information Criterion (Chao and Phillips, 1999)
KP: Kleibergen and Paap method with diffuse prior (2002)
Strachan: Strachan's method with diffuse prior, Strachan and Inder (2004)

LR: Johansen's LR trace test (Johansen, 1991)

Table 2: Model Selection and the number of the breaks m by the Posterior Probabilities

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	Pr(Model|Y)
Model 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model 3	0.0000	0.0004	0.0002	0.8357	0.1354	0.0000	0.9717
Model 4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model 5	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0002
Model 6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Model 7	0.0000	0.0005	0.0092	0.0185	0.0000	0.0000	0.0282
Model 8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr(m|Y)	0.0000	0.0011	0.0094	0.8542	0.1354	0.0000	

Note: m denotes the number of the breaks

Table 3: Estimates of the Break Points for Model 3

()=standard deviation,

	Post. Mode for Model 3	95% HPDI
d_1	1991:04 (0.2611)	1991:02, 1991:07
d_2	1999:04 (5.3813)	1998:08, 1999:08
d_3	2001:07 (1.4309)	2001:05, 2001.10

Figure 3: Posterior Probability Mass of the Break Points for Model 3

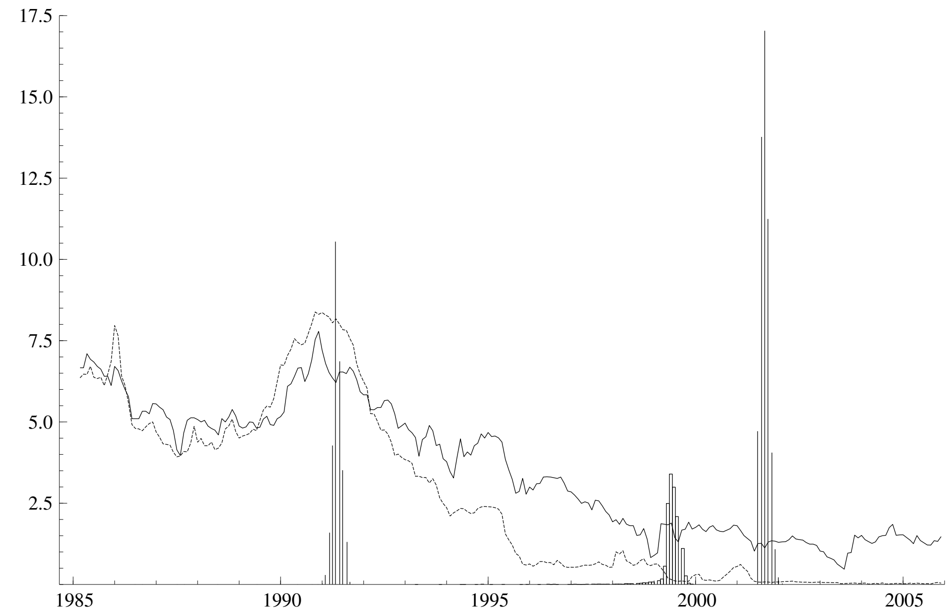

Table 4: Parameter Estimates (Posterior Mean) for Model 3

()=standard deviation,

Parameter	$r_{l,t}$	$r_{s,t}$
α_1	-0.00364 (0.00095)	0.00123 (0.00039)
α_2	-0.00512 (0.00065)	0.00506 (0.00123)
α_3	0.00085 (0.00086)	0.00154 (0.00153)
α_4	-0.00408 (0.00414)	0.00055 (0.00109)

δ_1	δ_2	δ_3	δ_4
-0.39609	1.6057	1.0528	1.3496
(0.09099)	(0.04563)	(0.05371)	(0.03094)

$$\Omega_1 = \begin{bmatrix} 0.05268 & 0.02764 \\ (0.02764) & (0.00632) \\ 0.02764 & 0.10592 \\ (0.00632) & (0.01958) \end{bmatrix}, \Omega_2 = \begin{bmatrix} 0.05278 & 0.01106 \\ (0.00762) & (0.00168) \\ 0.01106 & 0.02468 \\ (0.00168) & (0.00344) \end{bmatrix}$$

$$\Omega_3 = \begin{bmatrix} 0.01206 & 0.00099 \\ (0.01046) & (0.00063) \\ 0.00099 & 0.00374 \\ (0.00063) & (0.00149) \end{bmatrix}, \Omega_3 = \begin{bmatrix} 0.03651 & 0.00050 \\ (0.01092) & (0.00027) \\ 0.00050 & 0.00027 \\ (0.00027) & (0.00007) \end{bmatrix}$$

Figure 4: Posterior Density of α for Model 7