Cointegration and Regime-Switching Risk Premia in the U.S. Term Structure of Interest Rates

Peter Tillmann[1]
University of Bonn
Institute for International Economics
Lennéstr. 37, D-53113 Bonn
tillmann@iiw.uni-bonn.de

first version: July 2003, this version: May 2004

Abstract: To date the cointegrating properties and the regime-switching behavior of the term structure are two separate strands of the literature. This paper integrates these lines of research and introduces regime shifts into a cointegrated VAR model. We argue that the short-run dynamics of the cointegrated model are likely to shift across regimes while the equilibrium relation implied by the expectations hypothesis of the term structure is robust to regime shifts. A Markov-switching VECM approach for U.S. data outperforms a linear VECM. We find significant shifts in risk premia and interest rate volatility. These regime shifts reflect changing inflation expectations and shifts in monetary policy, respectively.

Keywords: term structure, expectations hypothesis, cointegration, Markov-switching, monetary policy

JEL classification: E43, E52

[1]Paper prepared for the 2004 North American Summer Meeting of the Econometric Society at Brown University. I thank Katrin Assenmacher-Wesche, Jens Clausen, Michael Massmann, and Manfred J.M. Neumann as well as participants of the Deutsche Bundesbank research seminar and the Halle Wokshop "Makroökonometrie" for helpful comments on an earlier draft. All remaining errors are mine.

1 Introduction

The information content of the term structure of interest rates has been studied intensively. Despite the poor empirical performance of the leading theoretical model, the expectations hypothesis, the yield curve is widely used as an indicator of monetary and financial conditions. According to this theory, the spread between long- and short-term yields contains information about the future course of interest rates. This paper sidesteps these short-run issues and focuses on the long-run implications of the expectations hypothesis. The hypothesis implies that long and short rates should be cointegrated with cointegrating coefficients summing to zero.

While the cointegration properties of the term structure are studied widely, another strand of multivariate modelling analyzes regime shifts in the stochastic processes generating interest rates. These lines of research are largely separate strands of the literature. Furthermore, recent research points to instability in the short-run dynamics of cointegrating models of the term structure. These studies either assume one-time structural shifts at predetermined dates or non-linearities governed by an observable threshold. Thus far the cointegration properties and the Markov-switching behavior have not been studied jointly. Previous cointegration studies are not capable to shed light on shifting risk premia and other regime-dependent dynamics which are likely to be induced by shifts in monetary policy.

This paper provides an unifying approach and introduces regime shifts into the cointegrated VAR model of the term structure. The state variable is unobservable and the model endogenously determines the characteristics of the regimes and the break dates. Drawing on recent empirical research this paper argues that the cointegrating relation linking long and short yields is likely to be robust to regime shifts while the short-run dynamics including the term premium and the equilibrium adjustment are dependent on the prevailing unobservable regime. Thus, this paper reconciles fluctuations in stationary risk premia and error-correction parameters with the long-run equilibrium relation implied by the expectations hypothesis. This approach offers valuable insights into how monetary policy regimes are reflected in term structure dynamics.

We fit a Markov-switching vector error-correction model (MS-VECM) to monthly U.S. data where the risk premium, the short-run drifts, and the loadings are regime-dependent. Given the one-to-one cointegrating relation between the three-months and various long rates and, thus, the stationarity of risk premia, the model is able to detect discrete shifts in the stochastic process corresponding to well known episodes of U.S. monetary policy.

The model identifies two distinct regimes that differ mostly with respect to interest rate volatility. We find that the high-variance regime prevails during the non-borrowed reserve-targeting episode of Federal Reserve policy in 1979-1982 and other periods of

rising inflation expectations. Shifts to this regime imply increasing risk premia at the short end of the term structure and decreasing risk premia for longer maturities. This is consistent with decreasing long-run inflation expectations accompanied by increasing inflation expectations for a short horizon up to twelve months. Furthermore, the adjustment of long rates towards the equilibrium yield spread is much faster when interest rate volatility is high. A second regime reflects the stable post-1987 period characterized by low premia for short and intermediate maturities, low volatility, and small expected changes in long-horizon interest rate forecasts. Thus, we supplement recent findings of e.g. Hansen (2003) who identifies regime shifts at predetermined dates. This paper, on the contrary, lets the model endogenously choose the dates of regime shifts and models recurrent structural change supported by a large literature as opposed to occasional structural instability.
The plan of the paper is the following: The next section derives the cointegrating properties from a simple exposition of the expectations hypothesis and provides a brief review of the literature. Section three sets up a linear VECM and tests the cointegrating properties for U.S. data while section four proposes a regime switching VECM approach and interprets the findings. Section five finally concludes.

2 Information in the term structure of interest rates

This section gives a brief overview of recent research on the equilibrium relationship between interest rates of different maturity. We first derive the cointegrating properties implied by a standard formulation of the expectations hypothesis of the term structure and then survey the existing evidence with a special focus on the regime-shifting behavior of interest rates and, hence, the term structure.

2.1 Cointegration and the expectations hypothesis

The expectations hypothesis of the term structure of interest rates implies a stable one-to-one relationship between short and long rates. Suppose an n-period pure discount bond yields $R_t(n)$ while the forward rate $F_t(n)$ is the yield from contracting at time t to buy a one period pure discount bond at time $t+n$ which matures at time $t+n+1$. Then it holds that $F_t(0) = R_t(1)$. The Fisher-Hicks formula gives

$$R_t(n) = \frac{1}{n} \sum_{j=0}^{n-1} F_t(j) \tag{1}$$

The expectations hypothesis says that

$$F_t(n) = E_t\left(R_{t+n}(1)\right) + \theta\left(n\right) \tag{2}$$

where $\theta(n)$ is the risk premium and E_t denotes the expectations operator based on information at time t. Substituting gives

$$\begin{aligned} R_t(n) &= \frac{1}{n}\left[\sum_{j=0}^{n-1}\left(E_t\left(R_{t+j}(1)\right)+\theta(j)\right)\right] \\ &= \frac{1}{n}\left[\sum_{j=0}^{n-1} E_t\left(R_{t+j}(1)\right)\right]+\phi(n) \end{aligned} \tag{3}$$

with $\phi(n) = \frac{1}{n}\sum_{j=0}^{n-1}\theta(j)$ as the average risk premium. The pure expectations hypothesis requires $\phi(n) = 0$, while weaker versions restrict this term to be constant or stationary. This no-arbitrage condition says that the long rate equals the weighted average of the expected short rates. The term premium measures the additional gain from holding long-term bonds relative to rolling-over one-period bonds. Using the identity

$$E_t(R_{t+j}(1)) = \sum_{i=1}^{j} E_t\left(\Delta R_{t+i}(1)\right) + R_t(1) \tag{4}$$

and rearranging results in

$$R_t(n) - R_t(1) = \frac{1}{n}\left[\sum_{j=1}^{n-1}\sum_{i=1}^{j} E_t\left(\Delta R_{t+i}(1)\right)\right] + \phi(n) \tag{5}$$

where Δ is the difference operator. Assuming that $R_t(1)$ and $R_t(n)$ are integrated of order one, $I(1)$, it follows that the right-hand side of (5) is stationary (given a stationary risk premium). Thus, the linear combination $R_t(n) - R_t(1)$ is stationary. In other words, the vector $x_t = [R_t(n), R_t(1)]'$ is cointegrated with a cointegration vector $\beta = (1, -1)'$. The necessary condition for the expectations condition to hold is that we can impose the restriction $\beta = (1, -1)'$ onto the yield spread. In this case the term premium is stationary.

The risk premium $\phi(n)$ will later be reflected as a constant in the cointegrating space. Note that the relation described by (5) holds for any pair $(n, 1)$. In the following we assume that the short rate is the three-months interest rate $R_t(3)$ and analyze the spread $R_t(n) - R_t(3)$ for $n \in \{6, 12, 36, 60, 120\}$ months.

The seminal work of Campbell and Shiller (1987) shows that present value models imply cointegration. They find a cointegrating vector of $(1, -1)'$ as required by the expectations hypothesis. These cointegrating properties of the term structure are also examined by Hall, Anderson, and Granger (1992) in a VECM for twelve variables. They find a cointegration vector consistent with the theory. Shea (1992) examines pairwise cointegration relations and finds mixed evidence. Engsted and Tanggaard

(1994) find support for the long-run implications of the expectations hypothesis for the U.S. while Cuthbertson (1996) provides support from UK interbank data.
Although these studies suggest that the term premium is stationary, a large body of research initiated by Engle, Lilien, and Robins (1987) confirms the time-varying nature of risk premia in excess holding yields that increase with volatility. Hence, a main point of term structure modelling is to quantify the size and the behavior of the term premium. This paper supplements existing cointegration studies by showing the dynamics of the term premium given its stationarity.

2.2 Regime shifts in the term structure

A large strand of the literature argues that regime shifts in monetary policy translate into regime shifts in interest rates and, thus, into regime-dependent behavior of the term structure. The change in the operating procedures of the Federal Reserve between 1979 and 1982 are frequently seen as a potential source of shifts in the term structure motivating many Markov-switching applications. In 1979 the Federal Reserve moved from interest rate targeting to money growth targeting and allowed the interest rate to fluctuate freely. This shift resulted in dramatically higher and more volatile short-term interest rates as can be seen from figure (1) and induced a change in the stochastic process of the entire term structure.
Sola and Driffill (1994) estimate a vector autoregression (VAR) for three and six months rates and allow for Markov regime shifts. They find their multivariate model to be more efficient than Hamilton's (1988) original regime-switching contribution. Regime shifts occur between 1979 and 1982 during the monetary targeting intermezzo of the Federal Reserve.[2] Similar studies by Kugler (1996) and Engsted and Nyholm (2000), among many others, for Swiss and Danish data support the regime-dependent behavior of the term structure and provide mixed results on the validity of the expectations hypothesis. The regime-dependent nature of term structure dynamics is a stylized fact.[3] However, the aforementioned studies model shifts in interest rates in a stationary VAR system in first differences since interest rates are likely to be I(1). Thus far the cointegrating properties and the regime shifts are treated separately.[4] This paper, on the contrary,

[2] Fuhrer (1996) shows that minor shifts in the coefficients of the central bank's reaction function can significantly affect the behavior and the information content of the term structure. Cogley (2003) uses a bivariate Bayesian VAR with interest rates of different maturity and allows for drifting conditional means and stochastic volatility of the innovation variance to study the changing nature of term structure dynamics.

[3] Additional evidence on the regime-dependent stochastic processes determining interest rates is presented in Ang and Bekaert (2002), Bekaert, Hodrick, and Marshall (2001), and Gray (1996).

[4] The short-term predictive power of the term structure for future interest rates may be severely impaired by the existence of a peso problem when the sample moments do not coincide with population moments taken into account by rational agents. Peso problems provide an additional motivation to

proposes a joint modelling approach.
Another line of research studies potential instability in cointegrated systems and applies various testing procedures to term structure data. Hansen (1992a) develops a Lagrange-Multiplier test for parameter instability and finds a stable one-to-one relationship. Hansen and Johansen (1999) elaborate a recursive maximum likelihood procedure that employs the time paths of the eigenvalues to analyze the stability of a VECM. This test confirms the constancy of the cointegrating vector for a set of four U.S. interest rates. Hansen (2003) generalizes Johansen's (1988) maximum likelihood procedure to allow for structural change. He finds significant changes in the short-run dynamics of the VECM in September 1979 and October 1982 but cannot reject the hypothesis of a stable long-run equilibrium. The risk premium, the variance-covariance matrix, and the adjustment coefficients are subject to discrete shifts while the cointegrating vector is unaffected by shifts in monetary policy. This econometric exercise, however, requires the dates of the regime shifts to be known in advance and tests for multiple breaks as compared to recurrent shifts between a predetermined number of distinct regimes. The attractiveness of the Markov-switching approach, on the contrary, is that the model endogenously separates regimes arising from a probabilistic process and dates their shifts without imposing a priori break dates.
Related studies argue that the term structure is characterized by non-linear and asymmetric adjustment towards the equilibrium in the sense that a regime-shift occurs once the spread exceeds a threshold. Hansen and Seo (2002) and Seo (2003) develop a threshold cointegration model and find evidence of non-linear mean reversion. While the state variable is observable in their case, this paper puts forward a regime-switching model with an unobservable state variable. Moreover, while these studies model non-linearity depending on the size and the sign of deviations from equilibrium, the model presented in this paper exhibits non-linearity over time.
It appears as a consensus view that the long-run cointegrating properties of the term structure are robust to regime shifts. In fact, Engsted and Tanggaard (1994, p. 175) argue that "the one-to-one relationship between long- and short-term rates given by the expectations hypothesis is not in any way dependent on the specific process generating short-term rates. If the expectations hypothesis is true, we therefore expect the cointegration implications to hold for the whole period and not just in periods of stable monetary policy". Hence, the low frequency properties of the term structure (i.e. the cointegrating vector) should be robust to regime shifts while the high frequency properties (i.e. the risk premium and the short-run dynamics) are likely to reflect regime shifts. This approach is pursued in the remainder of this paper.[5]

employ state-dependent regression models, see Bekaert, Hodrick, and Marshall (2001).

[5] In related work, Gutiérrez and Vázquez (2003) analyse how the predictive content of the spread for short rate changes has changed over the post-war period. They find one regime with a random-walk behavior of the short rate and another with a high and volatile rate where the spread has some

Recently, a strand of the finance literature incorporates regime-switching behavior in factor models of the term structure. Examples are Bansal, Tauchen, and Zhou (2003) or Dai, Singleton, and Yang (2003). These studies exclusively treat interest rates as stationary variables and, therefore, cannot account for long-run equilibrium relations. The studies surveyed here cannot detect the regime-switching dynamics of risk premia in the presence of cointegration in the long-run. As Kozicki and Tinsley (2002) note, there might be considerable variation in risk premia over time which are possibly related to the behavior of monetary authorities. We model a cointegrated VAR model for a pair of yields and allow for unobservable regime shifts in the term premium, the short-term drift, and the error-correction mechanism given a stable long-run equilibrium.

2.3 The data set

We employ interest rate series obtained from the Federal Reserve Bank of St. Louis.[6] The data set covers the period 1970:01 to 2004:02 at monthly frequency and comprises interest rates (in percent p.a.) on U.S. bonds for maturity (in months) $n \in \{3, 6, 12, 36, 60, 120\}$. A pairwise plot of the series is presented in figure (1). The spreads relative to the three-months rate are depicted in figure (2). Evidently, the mean of the interest rate spread as well as interest rate volatility experience structural changes across subperiods.

Standard Augmented Dickey-Fuller and Phillips-Perron tests, whose results are reported in table (2), cannot reject the hypothesis of a unit-root for each maturity. In other words, $R_t(n)$ is $I(1)$ as found by previous research.[7] Moreover, the Kwiatkowski-Phillips-Schmidt-Shin test rejects the hypothesis of stationarity. We refrain from testing for a unit-root in the interest rate spread since we will perform a much more powerful test for the prespecified cointegrating vector $\beta' = (1, -1)$ in the next section. Table (1) provides some descriptive statistics of interest rate differentials that will turn out to be useful for cross-checking subsequent results.

predictive ability.

[6] This data set is publicly available under `http://research.stlouisfed.org/fred2/`.

[7] While interest rates are found to be I(1) by the overwhelming majority of empirical research, it is certainly difficult to motivate this finding theoretically. An interesting perspective is offered by [1]Aït-Sahalia (1996). He tests different continous-time models for spot rates and finds that interest rates behave like a random-walk in the interval $[4\%, 17\%]$ while they show mean-reversion outside this spectrum. Therefore, it is difficult to reject the hypothesis of a unit-root over relatively short time periods.

3 The cointegrating properties in a linear VECM

To study the cointegrating and the regime-switching properties we proceed in two steps. In this section we develop a bivariate VECM approach for the term structure. The cointegrating properties are derived using Johansen's (1991) maximum likelihood procedure for a linear VECM. In a subsequent section the model is extended to include regime-dependent coefficients given these cointegrating properties.
Assume that we can describe the pairwise dynamics of long- and short-term interest rates by a bivariate VAR(q) system

$$x_t = v + A_1 x_{t-1} + ... + A_q x_{t-q} + \varepsilon_t \tag{6}$$

with $x_t = (R_t(n), R_t(3))'$ and normally distributed Gaussian innovations $\varepsilon_t \sim N(0, \Sigma)$. The intercept terms are collected in the (2×1) vector v. By subtracting x_{t-1} from both sides this system can be written as a vector error-correction model (VECM)

$$\Delta x_t = v + \Pi x_{t-1} + \sum_{i=1}^{q-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \tag{7}$$

with $\Pi = -\left(I - \sum_{i=1}^{q} A_i\right)$ and $\Gamma_i = -\left(A_{i+1} + ... + A_q\right)$ for $i = 1, ...q-1$.
Given that the variables in x_t are $I(1)$ Johansen (1991) formulates the hypothesis of cointegration as a reduced rank restriction on the Π matrix with

$$\Pi = \alpha\beta'$$

where α and β are $(2 \times r)$ matrices. We can interpret $rank(\Pi) = r$ as the number of stationary long-run relations while the cointegrating vector β is determined by solving an eigenvalue problem. Thus, $\beta' x_t$ is a stationary long-run equilibrium relation with the adjustment towards the equilibrium driven by the vector of loadings α.

Estimating the cointegrated VAR model requires a specification of a lag order q. The standard Akaike and Schwartz criteria as well as the Hannan-Quinn criterion reported in table (3) recommend different lag orders. Since subsequent models will be heavily parameterized we favour a parsimonious specification. We interpret the lag length recommended by the SC, which uses the higher penalty for extra coefficients and is therefore more rigid, as the guideline. However, the ultimate choice is determined following tests for serial correlation of the estimated residuals. Therefore, we estimate in VAR model with a lag length of $q = 4$.
The constant v is restricted to lie in the cointegrating space spanned by α for two reasons. First, a look at the data series in figure (1) does not suggest the presence of a linear trend. Second, the restricted constant in the cointegrating space corresponds to the risk premium derived in the theoretical discussion presented above.[8]

[8] The treatment of the constant has no effect on the estimated rank or the cointegrating properties.

The results of Johansen's (1991) maximum likelihood estimation of the $\Pi = \alpha\beta'$ matrix are presented in table (4). For each pair of maturities $x_t' = [R_t(n), R_t(3)]$ the trace test and the maximum eigenvalue test cannot reject the hypothesis of $r \leq 1$ while the hypothesis of $r = 0$ is clearly rejected in all cases. The strength of the cointegrating property weakens with maturity as reflected by the maximum eigenvalue $\lambda^{\max}$ which decreases as maturity n increases. Thus, we find strong evidence in favor of cointegration and can set $\hat{r} = 1$ in subsequent estimations.
In addition, the cointegration test developed by Horvath and Watson (1995) supports the presence of a cointegrating relation. Moreover, this test, which evaluates the hypothesis of no cointegration against the alternative of a cointegrating vector with unit coefficients, supports the prediction from the expectations hypothesis. Note that this test amounts to a standard Likelihood-Ratio test for the presence of the candidate error-correction terms in a first difference VAR and is more powerful than conventional unit-root tests applied to the interest rate spread.
To test the implications of the expectations hypothesis we impose restrictions onto the cointegrating vector $\beta = (\beta_{long}, \beta_{short})'$. In table (5) we normalize $\beta_{long} = 1$ and impose the restriction $\beta' = (1, -1)$ on the system. This restriction cannot be rejected in almost any of the scenarios using Likelihood Ratio tests. The restriction is rejected for the $[R_t(12), R_t(3)]$ pair. Given the results of the aforementioned powerful Horvath-Watson test, however, we proceed as if the unitary coefficient were accepted. Thus, we find strong support for the cointegrating implications of the expectations hypothesis: long and short rates cointegrate with a cointegrating vector $\beta' = (1, -1)$. The stationary linear combination indeed corresponds to the spread implied by the expectations hypothesis.
While at the short end of the term structure the long-run equilibrium is given by $R_t(6) - R_t(3) = 0.15$, the constant grows to 1.96 for the widest yield spread. As discussed earlier, this intercept in the cointegrating equation can be interpreted as a risk premium embedded in long rates. The risk premium increases monotonically with maturity.
The adjustment of Δx_t towards the long-run equilibrium is described by the vector of loading coefficients $\alpha = (\alpha_{long}, \alpha_{short})'$. In theory, both adjustment parameters should be positive because a larger spread $\beta' x_t$ means that long rates earn a higher interest rate, so long bonds must eventually depreciate and the long rate must rise to equilibrate the system. Since the expectations hypothesis claims that the long rate is an average of future short rates, the short rate is also expected to rise. However, we find this prediction to be satisfied only at the short end of the term structure, see table (5). In all other models, $\alpha_{long} < 0$ and $\alpha_{short} > 0$. This pattern is inconsistent with the short-run implications of the expectations hypothesis but is in line with the existing empirical evidence. Campbell and Shiller (1991, p. 496) argue: "In a nutshell, when

the spread is high the long rate tends to fall and the short rate tends to rise". Testing for weak exogeneity of either interest rate series amounts to restricting the respective adjustment coefficient to zero. In all but one case $\alpha_{short} = 0$ can be rejected while $\alpha_{long} = 0$ cannot be rejected at the short end only. Thus, some of the equilibrium adjustment occurs through movements in the short rate and not the long rate. This is somewhat surprising given that fact that movements in the short rate should closely reflect monetary policy set autonomously.

Before estimating the regime-switching model we test whether the residuals from the linear VECM exhibit non-linearity in the sense of deviation from the assumed IID distribution. For this purpose the Brock-Dechert-Scheinkman (BDS) diagnostic test is applied which tests the null hypothesis of linearity against an unspecified non-linear alternative.[9] The test statistic derived by Brock et al. (1996) is asymptotically normal and is reported in table (6) for alternative parameter constellations. For all VECM specifications the hypothesis of linearity is rejected at highest levels of significance. Thus, it seems that the linear VECM fails to capture non-linearities prevailing in the true data-generating process.

Since we cannot reject the long-run implications of the expectations hypothesis, we now turn to the analysis of regime shifts in the short-run dynamics given this estimated long-run equilibrium relationship.

4 A Markov-switching VECM

In this section a Markov-switching VECM is proposed that generalizes the model described by (7) to account for regime shifts. In other words, the model is piecewise linear in each regime but non-linear across regimes. If the number of regimes is set to unity, the model collapses to (7). Clarida et al. (2003) and Sarno, Thornton, and Valente (2002) use a similar approach, although for different purposes. They are primarily interested in the forecasting properties of the MS-VECM and do not disentangle regime-shifting parameters to gain information about the behavior of the term premia.

We model regime shifts given the one-to-one equilibrium relationship found in the previous section. Certainly, the well-established framework developed by Johansen (1991) models long-term properties for linear systems. However, recent work by Saikkonen and Luukkonen (1997) shows that these procedures originally developed for finite Gaussian VAR systems can be employed when the data are generated by an infinite non-Gaussian

[9] Psaradakis and Spagnolo (2002) compare the relative performance of portmanteau-type tests to detect nonlinearity generated by Markov regime-switching. They conclude that the BDS test is generally very powerful.

VAR.[10] Thus we follow the considerations of Krolzig (1997) and the empirical work by Krolzig, Marcellino, and Mizon (2002), Sarno, Thornton, and Valente (2002), Clarida et al. (2003), and Francis and Owyang (2003) and proceed in two steps by imposing the cointegrating properties derived in the linear model onto the regime-switching model.

4.1 Model specification

Suppose that the system describing short and long rates is driven by an unobservable discrete state variable $s_t = m$ with two possible regimes $m \in \{1, 2\}$

$$\begin{aligned} \Delta x_t &= v(s_t) + \Pi(s_t)\, x_{t-1} + \sum_{i=1}^{q-1} \Gamma_i(s_t)\, \Delta x_{t-i} + \varepsilon_t \\ \varepsilon_t &\sim N(0, \Sigma(s_t)) \end{aligned} \tag{8}$$

In contrast to the model in (7), the vector of intercept terms $v(s_t)$, the error-correction terms $\alpha(s_t)$, the dynamics of the stationary part $\Gamma_i(s_t)$, and the variance-covariance terms $\Sigma(s_t)$ of the innovations of this VECM are conditioned on the realization of the state variable. Note that β is regime-independent. Given the long-run equilibrium relationship we can safely impose $\beta' = (1, -1)$ derived in the previous section. Furthermore, we can decompose the regime-dependent vector of intercepts into one part entering the cointegrating space and one part affecting the short-run dynamics Δx_t

$$\Delta x_t - \delta(s_t) = \alpha(s_t)\left[\beta' x_{t-1} - \mu(s_t)\right] + \sum_{i=1}^{q-1} \Gamma_i(s_t)\left[\Delta x_{t-i} - \delta(s_t)\right] + \varepsilon_t \tag{9}$$

where

$$\begin{aligned} E(\Delta x_t | s_t) &= \delta(s_t) \\ E(\beta' x_{t-1} | s_t) &= \mu(s_t) \end{aligned}$$

with

$$\mu(s_t) = \left[\left(\alpha'\alpha\right)^{-1} \alpha'(\Gamma C - I) v | s_t\right] \tag{10}$$

and

$$\delta(s_t) = \left[\beta_\perp \left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1} \alpha_\perp v | s_t\right] \tag{11}$$

where $\Gamma = I - \Gamma_1 - ... - \Gamma_{q-1}$ and $C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha_\perp$. The orthogonal complements $\alpha_\perp$ and $\beta_\perp$ have full rank and are defined by $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. This decomposition is formally derived in the appendix. Thus, shifts in $v(s_t)$ translate into changes

[10] The power of residual-based cointegration tests, on the contrary, usually falls sharply in the presence of regime shifts. See Gregory and Hansen (1996) for this issue.

in the mean of the equilibrium relation $\mu(s_t)$ of the system and in the expected vector of short-run drifts $\delta(s_t)$. Hence, both Δx_t and $\beta' x_{t-1}$ are expressed as deviations from their means. In other words, each regime s_t is characterized by a particular attractor set $(\mu(s_t), \delta(s_t))$. Following Hansen (2003), among others, the coefficient μ corresponds to the term premium ϕ included in the theoretical model. Thus, the model is able to capture shifts in the risk premium μ along with shifts in the drift and in the variance-covariance matrix of the innovations. We relax the assumption of linear adjustment towards the equilibrium and let also the vector of adjustment coefficients $\alpha(s_t)$ and the matrices of the autoregressive part to be regime-dependent.
Hamilton (1988) proposes the application of unobservable Markov chains as regime-generating processes

$$prob(s_t = j|s_{t-1} = i, s_{t-2} = k, ...) = prob(s_t = j|s_{t-1} = i) = p_{ij} \tag{12}$$

In the class of models applied here the regime that prevails at time t is unobservable. The Markov property described in equation (12) says that the probability of a state m at time t, i.e. $s_t = m$, only depends on the state in the previous period, s_{t-1}. The transition probability p_{ij} says how likely state i will be followed by state j. Collecting the transition probabilities in a (2×2) matrix gives the transition matrix P

$$P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \tag{13}$$

where the element of the i-th column and the j-th row describes the transition probability $p_{ij.}$
Since the state variable is assumed to be unobservable, the estimation procedure is based on the iterative Baum-Lindgren-Hamilton-Kim-filter (BLHK-filter), that infers the regime-probabilities at each point in time.[11] As a by-product of the filter-inferences, a likelihood function is derived and maximized in order to obtain parameter estimates of model parameters. The log-likelihood function $L(\theta|Y_T)$ is given by the sum of the log-densities $f(.)$ of the observation y_t conditional on the history of the process $Y_t = \{y_\tau\}_{\tau=1}^{t}$ with a sample size T

$$L(\theta|Y_T) = \sum_{t=1}^{T} \ln f(y_t|Y_{t-1}; \theta) \tag{14}$$

with

$$\begin{aligned} f(y_t|Y_{t-1}; \theta) &= f(y_t, s_t = 1|Y_{t-1}; \theta) + f(y_t, s_t = 2|Y_{t-1}; \theta) \\ &= \sum_{m=1}^{2} f(y_t|s_t = m, Y_{t-1}; \theta) \cdot prob(s_t = m|Y_{t-1}; \theta) \end{aligned} \tag{15}$$

[11] Details about the estimation and filtering techniques are provided by Krolzig (1998).

where the second part of this expression follows from applying the rules of conditional probabilities saying that $f(y_t, s_t = m|...) = f(y_t|...) \cdot prob(s_t = m|...)$. The non-linear EM algorithm is applied to solve the problem

$$\widehat{\theta}_{ML} = \arg\max L(\theta|Y_T) \tag{16}$$

where the vector θ includes the MS-VECM-parameters to be estimated.

4.2 Results

The parameters of the Markov chain and some diagnostic tests are given in table (7). The maximum of the likelihood function obtained from the MS-VECM is substantially higher than that from the linear VECM. This max. $\ln L$ can be interpreted as a measure of the model's goodness of fit since the maximum likelihood estimator represents the value of the model's parameters for which the sample is most likely to have been observed. The Likelihood Ratio (LR) test under normal conditions does not apply here due to the existence of unidentified nuisance parameters (the transition probabilities are not identified under the linear model).[12] To circumvent this problem, a cautious approach is used. This implies that the LR test statistic is compared to a $\chi^2(d+n)$ distribution where d denotes the number of degrees of freedom and n stands for the number of nuisance parameters. Since the test statistic exceeds the critical value under this conservative benchmark, the null-hypothesis can be rejected at high significance levels. Although these test statistics must be interpreted somewhat cautiously, a non-linear regime switching specification seems to be superior to conventional linear models.
We restrict the analysis to a two-state Markov chain. Experiments with a three-state Markov chain in preliminary versions of this paper show that two of the resulting states are virtually indistinguishable from the high variance regime in the current paper. This choice is in line with the work of Ang and Bekaert (2002), Gutiérrez and Vázquez (2003), and others. Furthermore, we concentrate on regime shifts in all model parameters since regime-invariant autoregressive parameters, loadings, and covariances are rejected by Hansen (2003) and others.
The model endogenously separates distinct regimes characterized by regime-specific parameter sets. The estimated parameters of each regime are presented in table (8) and the regime-dependent attractor sets are reported in table (9). Several results are remarkable and provide a consistent picture:
First, regime 1 is characterized by a much higher variance of both the long and the short rate than regime 2. Thus, shifts in the underlying regime foremost affect the volatility

[12] The use of the standard χ^2 distribution would therefore cause a bias of the test against the null. See Hansen (1992b), Andrews and Ploberger (1994), and Garcia (1998) for this problem.

of interest rates.[13] Therefore, we subsequently characterize regimes primarily by as a high-variance regime (regime 1) and a low-variance regime (regime 2). In regime 1 (2) the variance of the short rate is higher (lower) than the variance of the long rate.
Second, the shifting vectors of adjustment coefficients α are in line with the interpretation of regimes put forward below. In state 1 the adjustment (in absolute terms) towards equilibrium is much stronger than in regime 2, which we will later interpret as a regime of stable monetary policy. Thus, interest rates adjust much faster in periods of unusual volatility which correspond to periods of rising inflation expectations and aggressive disinflation. Moreover, some of the estimated adjustment coefficients, especially those of the short rate in regime 2, are positive. The aforementioned discussion of the sign of the adjustment equally applies here. The negative sign of most coefficients reflects the empirical failure of the expectations hypothesis. We can conclude that in regime 1 the yield spread in period $t-1$ contains more information for the course of long rates in period t than in regime 2. We also find, as a general pattern, that the short rate contributes more to the error-correction mechanism in regime 2 while the long rate contributes more in regime 1. In state 1 only some of the α_{long} coefficients are significant (although with a negative sign) while in state 2 all α_{short} coefficients are significantly different from zero with a correct positive sign.
Third, as stressed by Hansen (2003), shifts in monetary policy have an important impact on the stochastic properties of interest rates and lead to substantial variation in risk premia. Disentangling the regime-dependent constant of the VECM into a regime-dependent mean of the equilibrium relation and a vector of drifts results in a regime-dependent risk premium given by $\mu(s_t)$. These equilibrium means μ and drift terms δ are presented in table (9). Risk premia mostly grow with maturity in each regime from $\mu = 0.12$ for regime 2 at the short end to $\mu = 1.92$ for the widest horizon. Regime 1 exhibits an higher risk premium at the short end of the term structure and a lower risk premium at the long end. The risk premium in state 1 peaks for a maturity of 12 months. This is consistent with the results provided by Hansen (2003). He finds (in what he calls model 2) that the regime prevailing between 1979 and 1982 leads to a higher risk premium for maturities below 12 months and a lower risk premium for longer maturities. These results are also consistent with the risk premia generated by the linear VECM ranging between 0.15 and 1.96. Moreover, a quick consistency check reveals that the regime-specific risk premia weighted by their unconditional regime probabilities documented in table (7) equal these numbers and also equal the mean spreads given in table (1).
Fourth, the high volatility regime exhibits high risk premia at the short end of the term

[13] Hansen (2003), among others, also finds evidence of structural breaks in the variance-covariance matrix of the VECM.

structure.[14] This is consistent with the finding of a time-varying term premium on long rates, see Engle Lilien, and Robins (1987), that has been proposed as an explanation for the failure of the expectations hypothesis to forecast interest rates. Furthermore, this lends support to the argument of Kozicki and Tinsley (2002) that more aggressive policy accompanied by a more volatile policy-controlled rate induces an upward shift in the term premium.

Fifth, regime 1 exhibits a mostly positive short-term drift while the drift in regime 2 is negative. Not surprisingly, the adjustment of interest rates is much stronger during periods of high volatility.

Sixth, the most remarkable regime shift occurs between 1979 and 1982, when the Federal Reserve changed its operating procedures. In this sense the results mirror the findings of other papers reviewed above. Figure (3) presents the conditional (smoothed) probabilities of regime 1 for each interest rate pair. During the 1979-82 period regime 1 prevails featuring high volatility. Regime 1 also reflects other phases of rising inflation expectations. Many of the dates of shifts to regime 1 correspond to the narrative account of Goodfriend (1993, 1998). He identifies periods of "inflation scare" accompanied by sharply rising long rates and decreasing anti-inflation credibility. However, the virtue of the regime-switching method is the ability to let the model detect regime shifts endogenously. According to the Fisher equation, $I(1)$ yields on long term bonds reflect long-run inflation expectations given a stationary real interest rate. Thus, shifts in risk premia correspond to shifts in inflation expectations. Between 1979 and 1982 risk premia at the short end rise while those at the long end fall. This indicates that long run inflation expectations decrease due to aggressive counter-inflation policy expressed in sharply rising short rates while inflation expectations rise over the short horizon. Following periods of persistent inflation expectations during the early and the late 1970s, the Fed under chairman Volcker engaged in aggressive disinflation policy. However, according to Goodfriend (1998), inflation expectations rose again in 1984. This re-emergence of inflation scare is reflected in the shift towards regime 1. The regime shifts occurring in 1973/74 and 1984 are also found, among others, by Ang and Bekaert (2002).[15] In regime 2, the volatility of both interest rates is low. Therefore, state 2 reflects monetary stability. Regime 2 coincides with the chairmanship of Alan Greenspan since 1987, indicating persistent anti-inflation credibility and a stable monetary environment.

Seventh, to facilitate the interpretation of the resulting regime-dependent adjustment coefficients and drift terms, we look at the common trends-implication of the coin-

[14] Cogley (2003) also finds that the risk premium on long-term bonds is connected to the variance of the short rate. He interprets this result as evidence of the connection between uncertainty about the target of monetary policy and mean yield spreads.

[15] The shift to regime 1 in 1973 might also reflect tensions in U.S. bond markets during the final breakup of the Bretton Woods system of fixed exchange rates.

tegrating relationship. Note that this analysis is purely illustrative as the notion of long-horizon forecasts is difficult to reconcile with stochastically switching parameters. If the variables are cointegrated, they share a common stochastic trend z_t, sometimes referred to as the permanent component, which can be viewed as the long-horizon forecast of the short-term interest rate

$$z_t = \lim_{j \to \infty} E_t \{R_{t+j}\} \tag{17}$$

Following King and Kurman (2002, p. 67), we interpret the stochastic trend as "describing permanent changes in the level of the short rate, which are reflected one-for-one in the long rate". Hence, changes in the properties of permanent component signal alternative stances of monetary policy or shifts in the perceived policy target for inflation.[16] Gonzalo and Granger (1995) propose the expression $\alpha'_{\perp} x_t$ as a simple representation of the common trend.[17] Thus, allowing for shifts in α implies that the stochastic trend in the interest rate system changes. It follows that expected changes of the trend are given by

$$E\left[\Delta\alpha'_{\perp} x_t | s_t\right] = \alpha'_{\perp}(s_t) E\left[\Delta x_t | s_t\right] = \alpha'_{\perp}(s_t)\delta\left(s_t\right) \tag{18}$$

The expected changes of the stochastic trend in each regime are presented in table (9). We find that the long-run interest rate level is expected to fall in regime 1 (except for the 12 months horizon). This is consistent with gaining long-run anti-inflation credibility in regime 1. Long-run interest rate forecasts in state 2, on the contrary, exhibit a moderate tendency to rise.

To summarize, we find that the term structure is subject to structural shifts induced by monetary policy. A shift to regime 1 increases volatility and strengthens the adjustment of long rates towards the equilibrium yield spread. Furthermore, risk premia at the short end of the yield curve are higher in regime 1 than in regime 2. Regime 1 prevails during the 1979-1982 episode and other periods of rising inflation expectations. Regime 2 reflects the stability of monetary policy in the post-1987 period.

5 Conclusions

It is widely argued that the stochastic process of interest rates is subject to discrete regime shifts. At the same time, the long-run implications of the term structure of interest rates are studied using exclusively linear, that is, regime-invariant models. This

[16] Kozicki and Tinsley (2001) provide a more detailed investigation of the implications of shifting policy targets in the sense of changes in limiting conditional interest rate forecasts for the term structure.

[17] Gonzalo and Granger (1995) refer to a common-factor representation since the permanent part is represented by a linear combination of observables.

paper argues that we can gain additional insights about the behavior of interest rates and the shifts in monetary policy by studying these two issues jointly. In particular, the regime-switching dynamics of stationary term premia can only be studied in a generalization of the cointegrated VAR model that allows for regime shifts.
We employed a Markov-switching VECM approach to analyze the behavior of the U.S. term structure given that interest rates of different maturity share a common stochastic trend. While the long-run equilibrium relation implied by the expectations hypothesis is likely to be stable over time, the short-run adjustment of interest rates towards the equilibrium as well as the term premium embedded in long rates shift between unobservable regimes governed by a first order Markov chain.
In accordance to the literature, we found these regime shifts to closely mirror the stance and the strategy of monetary policy. During the 1979-82 shift of the Federal Reserve from interest rate targeting to money growth targeting and other phases of inflation scare a regime prevails that exhibits a much higher variance and a much faster equilibrium adjustment than in the alternative regime. The risk premium at the short end increases while risk premia on long bonds decrease in this regime. This means that monetary policy leads to rising short-run inflation expectations but falling long-run inflation expectations. A regime with remarkable stability in terms of risk premia and interest rate volatility prevails in the post-1987 period. Thus, this paper contributed to closing the gap between two rather separate strands of the literature and, at the same time, provided evidence on the information content of the term structure over time.

6 Appendix: Decomposing the VECM constant

Consider the following N-dimensional VAR(q) in error-correction form where we drop the regime-dependence for convenience

$$\Delta x_t = v + \alpha\beta' x_{t-1} + \sum_{i=1}^{q-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t$$

where α and β are $(N \times r)$ matrices and $r = rank(\alpha\beta')$. The $(N \times 1)$ vector of unconstrained constants v can always be decomposed into two new vectors μ and δ such that one of them belongs to the cointegration space determined by α, $\mu \in sp(\alpha)$, and the other to Δx_t, $\delta \in sp(\beta_\perp)$.

Hansen and Johansen (1998, p. 31) show that the mean μ of the stationary process $\beta' x_t$ is given by

$$E\left(\beta' x_t\right) = \mu = \left(\alpha'\alpha\right)^{-1} \alpha'(\Gamma C - I)v$$

where $\Gamma = I - \Gamma_1 - ... - \Gamma_{q-1}$ and $C = \beta_\perp \left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1} \alpha_\perp$. To derive this expression, we follow their line of reasoning here: The Granger representation theorem (see Johansen (1995), p. 49) implies that the $I(1)$ series can be written as

$$x_t = C \sum_{i=1}^{t} (\varepsilon_i + v) + \sum_{i=0}^{\infty} C_i(\varepsilon_{t-i} + v) + A$$

where C_i are the moving average coefficient matrices and A depends on initial values such that $\beta' A = 0$. This implies that $E(\Delta x_t) = Cv$. Hence, taking expectations of the VAR(q) process gives the vector of means of the stationary series Δx_t

$$\begin{aligned} E(\Delta x_t) &= Cv = \alpha E\left(\beta' x_t\right) + \sum_{i=1}^{q-1} \Gamma_i C v + v \\ &= \beta_\perp \left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1} \alpha_\perp v \\ &\equiv \delta \end{aligned}$$

where $E\left(\beta' x_t\right) = E\left(\beta' x_{t-1}\right)$ holds due to the stationarity of $\beta' x_t$. With $q = 1$ this simplifies to

$$\delta = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha_\perp v$$

Multiplying $E(\Delta x_t)$ by $(\alpha'\alpha)^{-1} \alpha'$ and rearranging gives

$$\begin{aligned} E\left(\beta' x_t\right) &= \left(\alpha'\alpha\right)^{-1} \alpha' \left(I - \sum_{i=1}^{q-1} \Gamma_i\right) Cv - \left(\alpha'\alpha\right)^{-1} \alpha' v \\ &= \left(\alpha'\alpha\right)^{-1} \alpha'(\Gamma C - I)v \\ &\equiv \mu \end{aligned}$$

If $q = 1$, this reduces to

$$\begin{aligned} E\left(\beta' x_t\right) &= \left(\alpha'\alpha\right)^{-1}\alpha'(\Gamma C - I)v \\ &= \left(\alpha'\alpha\right)^{-1}\alpha'\left[\beta_\perp\left(\alpha'_\perp\Gamma\beta_\perp\right)^{-1}\alpha_\perp - I\right]v \\ &= \left(\alpha'\alpha\right)^{-1}\alpha'\left[-\alpha\left(\beta'\alpha\right)^{-1}\beta'\right]v \\ &= -\left(\beta'\alpha\right)^{-1}\beta' v \end{aligned}$$

Note that the vectors of equilibrium means and of drift terms are related by

$$v = \Gamma\delta - \alpha\mu$$

Let us now re-introduce regime-dependent coefficients $\alpha(s_t), v(s_t)$, and $\Gamma_i(s_t)$ as described in the text. The MS-VECM enables us to decompose the regime-dependent constant $v(s_t)$ into regime-specific attractor sets $(\mu(s_t), \delta(s_t))$ as explained in the text. This means there are $(N-r)$ linearly independent but state-dependent drifts collected in δ and r linearly independent but state-dependent equilibrium means collected in μ

$$\begin{aligned} E\left(\Delta x_t | s_t\right) &= \delta(s_t) \\ E\left(\beta' x_{t-1} | s_t\right) &= \mu(s_t) \end{aligned}$$

Hence, both Δx_t and $\beta' x_{t-1}$ are expressed as deviations from their regime-specific means.

References

[1] Aït-Sahalia, Y. (1996): "Testing Continous-Time Models of the Spot Interest Rate", *The Review of Financial Studies* 9, 385-426.

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

[3] Ang, A. and G. Bekaert (2002): "Regime Switches in Interest Rates", *Journal of Business and Economic Statistics* 20, 163-182.

[4] Bansal, R., G. Tauchen, and H. Zhou (2003): "Regime-Shifts, Risk Premiums in the Term Structure, and the Business Cycle", *unpublished*, Duke University.

[5] Bekaert, G., R. J. Hodrick, and D. A. Marshall (2001): "Peso problem explanations for term structure anomalies", *Journal of Monetary Economics* 48, 241-270.

[6] Brock, W. A., W. D. Dechert, J. A. Scheinkman, and B. LeBaron (1996): "A test for independence based on the correlation dimension", *Econometric Reviews* 15, 197-235.

[7] Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models", *Journal of Political Economy* 95, 1062-1088.

[8] Campbell, J. Y. and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View", *Review of Economic Studies* 58, 495-514.

[9] Clarida, R. H., L. Sarno, M. P. Taylor, and G. Valente (2003): "The out-of-sample success of term structure models as exchange rate predictors: a step beyond", *Journal of International Economics* 60, 61-83.

[10] Cogley, T. (2003): "An Exploration of Evolving Term Structure Relations", *unpublished*, University of California, Davis.

[11] Cuthbertson, K. (1996): "The Expectations Hypothesis of the Term Structure: The UK Interbank Market", *The Economic Journal* 106, 578-592.

[12] Dai, Q., K. J. Singleton, and W. Yang (2003): "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields", *unpublished*, Stern School of Business.

[13] Engle, R. F., D. M. Lilien, and R. P. Robins (1987): Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica* 55, 391-407.

[14] Engsted, T. and K. Nyholm (2000): "Regime shifts in the Danish term structure of interest rates", *Empirical Economics* 25, 1-13.

[15] Engsted, T. and C. Tanggaard (1994): "Cointegration and the US Term Structure", *Journal of Banking and Finance* 18, 167-181.

[16] Francis, N. and M. T. Owyang (2003): "Monetary Policy in a Markov-Switching VECM: Implications for the Cost of Disinflation and the Price Puzzle", *unpublished*, Federal Reserve Bank of St. Louis.

[17] Fuhrer, J. C. (1996): "Monetary Policy Shifts and Long-Term Interest Rates", *Quarterly Journal of Economics* 111, 1183-1209.

[18] Garcia, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review* 39, 763-788.

[19] Gonzalo, J. and C. Granger (1995): "Estimation of Common Long-Memory Components in Cointegrated Systems", *Journal of Business and Economic Statistics* 13, 27-35.

[20] Goodfriend, M. (1993): "Interest Rate Policy and the Inflation Scare Problem: 1979-1992", *Economic Quarterly* 79, Federal Reserve Bank of Richmond, 1-23.

[21] Goodfriend, M. (1998): "Using the Term Structure of Interest Rates for Monetary Policy", *Economic Quarterly* 84, Federal Reserve Bank of Richmond, 13-30.

[22] Gray, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process", *Journal of Financial Economics* 42, 27-62.

[23] Gregory, A. W. and B. E. Hansen (1996): "Residual-based tests for cointegration in models with regime shifts", *Journal of Econometrics* 70, 99-126.

[24] Gutiérrez, M.-J. and J. Vázquez (2003): "The changing behavior of the term structure of post-war U.S. interest rates", *unpublished*, Universidad del País Vasco.

[25] Hall, A., D., H. M. Anderson, and C. W. J. Granger (1992): "A Cointegration Analysis of Treasury Bill Yields", *The Review of Economics and Statistics* 74, 116-126.

[26] Hamilton, J. D. (1988): "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates", *Journal of Economic Dynamics and Control* 12, 385-423.

[27] Hansen, B. E. (1992a): "Tests for Parameter Instability in Regressions with I(1) Processes", *Journal of Business and Economic Statistics* 10, 321-335.

[28] Hansen, B. E. (1992b): "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics* 7, S61-S82.

[29] Hansen, B. E. and B. Seo (2002): "Testing for two-regime threshold cointegration in vector error-correction models", *Journal of Econometrics* 110, 293-318.

[30] Hansen, H. and S. Johansen (1998): *Workbook on Cointegration*, Oxford: Oxford University Press.

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

[32] Hansen, P. R. (2003): "Structural Changes in the Cointegrated Vector Autoregressive Model", *Journal of Econometrics* 114, 261-295.

[33] Horvath, M. T. K. and M. W. Watson (1995): "Testing for cointegration when some of the cointegrating vectors are prespecified", *Econometric Theory* 11, 984-1014.

[34] Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica* 59, 1551-1580.

[35] King, R. G. and A. Kurmann (2002): "Expectations and the Term Structure of Interest Rates: Evidence and Implications", *Economic Quarterly* 88, Federal Reserve Bank of Richmond, 49-95.

[36] Kozicki, S. and P. A. Tinsley (2001): "Shifting endpoints in the term structure of interest rates", *Journal of Monetary Economics* 47, 613-652.

[37] Kozicki, S. and P. A. Tinsley (2002): "Term Premia: Endogenous Constraints on Monetary Policy", *Working Paper*, No. 02-07, Federal Reserve Bank of Kansas City.

[38] Krolzig, H.-M. (1997): "Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts", *unpublished,* Nuffield College, Oxford.

[39] Krolzig, H.-M. (1998), "Econometric Modeling of Markov-Switching Vector Autoregressions using MSVAR for Ox", *unpublished*, Nuffield College, Oxford.

[40] Krolzig, H.-M., M. Marcellino, and G. E. Mizon (2002): "A Markov-switching vector equilibrium correction model of the UK labour market", *Empirical Economics* 27, 233-254.

[41] Kugler, P. (1996): "The term structure of interest rates and regime shifts: Some empirical results", *Economics Letters* 50, 121-126.

[42] Osterwald-Lenum, M. (1992): "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics", *Oxford Bulletin of Economics and Statistics* 54, 461-472.

[43] Psaradakis, Z. and N. Spagnolo (2002): "Power Properties of Nonlinearity Tests for Time Series with Markov Regimes", *Studies in Nonlinear Dynamics and Econometrics* 6, article 2.

[44] Saikkonen, P. and R. Luukkonen (1997): "Testing cointegration in infinite order vector autoregressive processes", *Journal of Econometrics* 81, 93-126.

[45] Sarno, L., D. L. Thornton, and G. Valente (2002): "Federal Funds Rate Prediction", *Working Paper*, No. 2002-005B, Federal Reserve Bank of St. Louis.

[46] Seo, B. (2003): "Nonlinear mean reversion in the term structure of interest rates", *Journal of Economic Dynamics and Control* 27, 2243-2265.

[47] Shea, G. S. (1992): "Benchmarking the Expectations Hypothesis of the Interest-Rate Term Structure: An Analysis of Cointegration Vectors", *Journal of Business and Economic Statistics* 10, 347-366.

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

Table 1: Descriptive statistics of interest rate spreads

spread (in %)	mean	max.	min.	std.dev.	skew	kurt	# obs.
$R_t(6) - R_t(3)$	0.135	1.450	-1.010	0.222	-0.069	10.812	410
$R_t(12) - R_t(3)$	0.221	2.375	-1.991	0.455	-0.194	6.568	410
$R_t(36) - R_t(3)$	1.145	4.110	-2.010	0.868	-0.454	3.881	410
$R_t(60) - R_t(3)$	1.390	4.330	-2.250	1.072	-0.596	3.358	410
$R_t(120) - R_t(3)$	1.647	4.420	-2.650	1.308	-0.604	2.995	410

Table 2: Unit-root tests

series	specification	ADF	PP	KPSS	ADF(4)	PP(4)	KPSS(4)
$R_t(3)$	const.	-1.17	-1.51	0.85***	-1.75	-1.87	2.58***
	no const.	-1.03	-1.29		-1.16	-1.27	
$R_t(6)$	const.	-1.49	-1.71	0.91***	-1.58	-1.71	2.78***
	no const.	-1.02	-1.22		-1.12	-1.21	
$R_t(12)$	const.	-1.07	-1.67	0.92***	-1.46	-1.64	2.84***
	no const.	-0.98	-1.18		-1.11	-1.16	
$R_t(36)$	const.	-1.26	-1.43	0.93***	-1.28	-1.41	2.91***
	no const.	-0.91	-1.02		-0.99	-1.04	
$R_t(60)$	const.	-1.21	-1.39	0.93***	-1.25	-1.35	2.91***
	no const.	-0.84	-0.96		-0.92	-0.96	
$R_t(120)$	const.	-1.13	-1.26	0.89***	-1.20	-1.23	2.84***
	no const.	-0.73	-0.82		-0.83	-0.82	

Notes: Unit-root tests with and without intercept term. ADF denotes the test statistic from the augmented Dickey-Fuller test, PP denotes the test statistic from the Phillips-Perron test, and KPSS is the Kwiatkowski-Phillips-Schmidt-Shin test statistic. While ADF and PP test the hypothesis of a unit-root, KPSS tests the Null of stationarity against the unit-root hypothesis. The lag order for the ADF test is chosen according to the Schwartz criterion; the PP and the KPSS test are specified using the Bartlett kernel with automatic Newey-West bandwidth selection. The last three columns report test results with a lag length set to four. A significance level of 1%, 5%, and 10% is indicated by ***, **, and *.

Table 3: Choosing the lag order of the VAR system

x'_t		AIC(q)	SC(q)	HQ(q)	LM(1)	LM(4)
$[R_t(6), R_t(3)]$	$q = 1$	0.196	0.256	0.220	73.38***	3.27
	$q = 2$	0.031	0.130	0.070	26.88***	3.36
	$q = 3$	-0.014	0.125	0.041	7.79*	5.22
	$q = 4$	-0.012	0.167	0.059	0.57	2.47
	$q = 5$	0.004	0.222	0.090	1.97	12.01**
$[R_t(12), R_t(3)]$	$q = 1$	1.408	1.468	1.432	91.35***	3.79
	$q = 2$	1.200	1.299	1.239	38.40***	3.31
	$q = 3$	1.124	1.263	1.179	12.04**	12.57**
	$q = 4$	1.118	1.297	1.189	4.09	7.96*
	$q = 5$	1.119	1.338	1.206	7.46	23.16***
$[R_t(36), R_t(3)]$	$q = 1$	1.898	1.958	1.922	101.85***	1.61
	$q = 2$	1.664	1.763	1.703	37.08***	0.57
	$q = 3$	1.588	1.727	1.643	17.52***	5.33
	$q = 4$	1.569	1.748	1.640	5.48	6.22
	$q = 5$	1.570	1.789	1.657	7.38	18.09***
$[R_t(60), R_t(3)]$	$q = 1$	1.889	1.948	1.912	98.66***	1.45
	$q = 2$	1.661	1.760	1.700	41.56***	0.38
	$q = 3$	1.574	1.713	1.629	20.65***	4.85
	$q = 4$	1.549	1.728	1.620	7.46	8.42*
	$q = 5$	1.549	1.764	1.632	6.74	19.65***
$[R_t(120), R_t(3)]$	$q = 1$	1.768	1.827	1.791	99.41***	1.01
	$q = 2$	1.542	1.641	1.581	36.86***	0.21
	$q = 3$	1.467	1.606	1.522	13.58***	2.98
	$q = 4$	1.452	1.631	1.523	5.40	5.27
	$q = 5$	1.454	1.673	1.541	5.15	15.59***

Notes: AIC(q), SC(q), and HQ(q) denote the Akaike information criterion, the Schwartz criterion, and the Hannan-Quinn information criterion, respectively, for a VAR of order q estimated in levels. These criteria compare the goodness of the fit of maximum likelihood estimations and correct for the loss of degrees of freedom when additional lags are added. LM(h) is a multivariate Lagrange-Multiplier test for residual correlation up to order h. Under the null hypothesis of no serial correlation of order h, the LM statistic is asymptotically χ^2 distributed with 4 degrees of freedom. A significance level of 1%, 5%, and 10% is indicated by ***, **, and *.

Table 4: Results of cointegration tests

(a) Johansen test

x'_t	H_0 $rank = r$	λ^{max}	trace test statistic	trace test 5% cv	λ^{max} test statistic	λ^{max} test 5% cv
$[R_t(6), R_t(3)]$	$r = 0$	0.08	36.87	19.96	34.25	15.67
	$r \leq 1$	0.01	2.62	9.24	2.62	9.24
$[R_t(12), R_t(3)]$	$r = 0$	0.08	37.21	19.96	34.49	15.67
	$r \leq 1$	0.01	2.72	9.24	2.72	9.24
$[R_t(36), R_t(3)]$	$r = 0$	0.07	30.44	19.96	27.77	15.67
	$r \leq 1$	0.01	2.67	9.24	2.67	9.24
$[R_t(60), R_t(3)]$	$r = 0$	0.06	27.95	19.96	25.27	15.67
	$r \leq 1$	0.01	2.68	9.24	2.68	9.24
$[R_t(120), R_t(3)]$	$r = 0$	0.05	23.71	19.96	21.40	15.67
	$r \leq 1$	0.01	2.32	9.24	2.32	9.24

(b) Horvath-Watson test

x'_t	test statistic	5% cv
$[R_t(6), R_t(3)]$	32.17	10.18
$[R_t(12), R_t(3)]$	32.44	10.18
$[R_t(36), R_t(3)]$	25.64	10.18
$[R_t(60), R_t(3)]$	23.09	10.18
$[R_t(120), R_t(3)]$	19.09	10.18

Notes: Johansen test for VECM with three lags (in differences) and a constant restricted to the cointegrating space. λ^{max} denotes the maximum eigenvalue. The trace test and the λ^{max} test are explained in detail in Johansen (1991). The 5% critical values are from Osterwald-Lenum (1992), table 1. The Horvath-Watson test of the null hypothesis of no cointegration against the known alternative of rank $r = 1$ with $\beta' = (1, -1)$ corresponds to a Wald test for the inclusion of error-correction terms, i.e. the interest rate spread, in a VAR in first differences with an unrestricted constant. Its test statistic is computed as $2(\ln L_{VECM} - \ln L_{VAR})$, where L denotes the value of the likelihood function under the respective model. The critical values are from Horvath-Watson (1995), table 1. Critical values for the case of a restricted constant are not yet available.

Table 5: Identification of the cointegrating space

	$R_t(6)$ $R_t(3)$	$R_t(12)$ $R_t(3)$	$R_t(36)$ $R_t(3)$	$R_t(60)$ $R_t(3)$	$R_t(120)$ $R_t(3)$
const.	0.15 (0.07)	0.20 (0.17)	1.02 (0.38)	1.52 (0.47)	1.96 (0.06)
β_{short}	-0.99 (0.01)	-1.07 (0.02)	-1.01 (0.05)	-0.97 (0.07)	-0.94 (0.09)
$H_0 : \beta_{short} = -1$					
LR (χ^2)	0.04	7.33	0.06	0.11	0.26
p	0.84	0.01	0.80	0.73	0.61
α_{long}	0.13 (0.12)	-0.01 (0.06)	-0.04 (0.02)	-0.04 (0.02)	-0.03 (0.01)
α_{short}	0.32 (0.13)	0.14 (0.06)	0.05 (0.03)	0.04 (0.02)	0.04 (0.02)
$H_0 : \alpha_{long} = 0$					
LR (χ^2)	1.07	0.07	3.31	5.31	5.24
p	0.30	0.80	0.07	0.02	0.02
$H_0 : \alpha_{short} = 0$					
LR (χ^2)	5.82	5.18	3.06	2.42	2.80
p	0.01	0.02	0.08	0.19	0.09

Notes: The constant is restricted to the cointegrating space. Standard errors in parenthesis. The cointegrating vector is normalized on the long rate, $\beta' = [1,\ \beta_{short}]$, the vector of adjustment coefficient is $\alpha' = [\alpha_{long},\ \alpha_{short}]$. The Likelihood Ratio (LR) test statistic of the hypothesis $\beta_{short} = -1$, i.e. the cointegrating vector $(1, -1)'$, and of the hypothesis of weakly exogenous long or short rates, i.e. $\alpha_i = 0$, is asymptotically χ^2 distributed. The marginal significance level is given by p.

Table 6: Linearity tests on VECM residuals

BDS specification		$\frac{R_t(6)}{R_t(3)}$	$\frac{R_t(12)}{R_t(3)}$	$\frac{R_t(36)}{R_t(3)}$	$\frac{R_t(60)}{R_t(3)}$	$\frac{R_t(120)}{R_t(3)}$
$w = 2$	$\eta = 0.5\sigma$	0.04***	0.02***	0.01***	0.01***	0.00***
		0.04***	0.04***	0.03***	0.03***	0.03***
	$\eta = 1.5\sigma$	0.05***	0.04***	0.02***	0.02***	0.02***
		0.06***	0.05***	0.05***	0.05***	0.10***
$w = 3$	$\eta = 0.5\sigma$	0.04***	0.03***	0.01***	0.05***	0.04***
		0.04***	0.05***	0.04***	0.04***	0.04***
	$\eta = 1.5\sigma$	0.10***	0.08***	0.04***	0.04***	0.04***
		0.11***	0.10***	0.10***	0.10***	0.10***

Notes: BDS test for iid-linearity against an unspecified alternative applied to the residuals from the linear VECM. The test statistic is asymptotically normal. The distance parameter is given by η, which is set equal to 0.5 and 1.5 times the standard deviation σ as recommended by Brock et al. (1996). The maximum dimension is given by w. A significance level of 1% is indicated by ***. Bootstrapped p-values indicate virtually identical levels of significance and are not reported here.

Table 7: Results from MS-VECM estimations

model	$R_t(6)$ $R_t(3)$	$R_t(12)$ $R_t(3)$	$R_t(36)$ $R_t(3)$	$R_t(60)$ $R_t(3)$	$R_t(120)$ $R_t(3)$
linear VECM					
max. ln L	22.03	-211.31	-298.01	-293.82	-275.71
AIC	-0.015	1.135	1.562	1.576	1.452
MSIAH-VECM					
max. ln L	310.49	96.34	-37.68	-45.67	-39.49
LR	576.93	615.31	520.65	496.29	472.43
$p\left[\chi^2(21)\right]$	0.00	0.00	0.00	0.00	0.00
AIC	-1.332	-0.278	0.383	0.458	0.392
p_{11}	0.92	0.88	0.91	0.92	0.90
p_{22}	0.98	0.96	0.98	0.98	0.98
$dur(1)$	12.11	8.30	11.55	12.79	10.36
$dur(2)$	51.44	23.15	53.42	62.11	43.01
$prob(1)$	0.19	0.26	0.18	0.17	0.19
$prob(2)$	0.81	0.74	0.82	0.83	0.81

Notes: The Likelihood Ratio (LR) test of the linear model against the Markov-switching alternative is computed as $LR = 2(\ln L_{MS-VECM} - \ln L_{VECM})$ where L denotes the value of the likelihood function under the respective model. The test statistic is χ^2 distributed with the degrees of freedom corrected for unidentified nuisance as explained in the text. The marginal significance level is denoted by p. AIC is the complexity-penalizing Akaike information criterion. The transition probabilities obtained from the Markov-switching model are denoted by p_{ii}, the expected duration (in months) of each regime s_t is denoted by $dur\,(s_t)$, and $prob(s_t)$ is the unconditional probability of each regime s_t.

Table 8: Parameter estimates from the MSIAH-VECM

	$R_t(6)$ $R_t(3)$	$R_t(12)$ $R_t(3)$	$R_t(36)$ $R_t(3)$	$R_t(60)$ $R_t(3)$	$R_t(120)$ $R_t(3)$
$v\,(1)$	0.027 (0.10) −0.014 (0.11)	0.206 (0.13) 0.001 (0.14)	0.101 (0.09) −0.020 (0.13)	0.097 (0.07) −0.028 (0.13)	0.089 (0.06) −0.034 (0.12)
$v\,(2)$	−0.023 (0.02) −0.039 (0.02)	−0.030 (0.03) −0.080 (0.02)	0.003 (0.04) −0.066 (0.03)	0.012 (0.03) −0.055 (0.02)	0.004 (0.03) −0.062 (0.02)
$\alpha(1)$	0.120 (0.34) 0.351 (0.38)	−0.236 (0.14) −0.030 (0.15)	−0.141 (0.07) 0.036 (0.10)	−0.134 (0.05) 0.052 (0.10)	−0.103 (0.04) 0.065 (0.08)
$\alpha(2)$	0.099 (0.10) 0.239 (0.09)	0.032 (0.05) 0.135 (0.04)	−0.010 (0.03) 0.049 (0.02)	−0.014 (0.02) 0.033 (0.01)	−0.007 (0.01) 0.030 (0.01)
$\tilde{\Sigma}\,(1)$	0.874 0.960	0.760 0.836	0.640 0.959	0.543 0.975	0.453 0.942
$\tilde{\Sigma}\,(2)$	0.232 0.211	0.233 0.170	0.276 0.208	0.259 0.213	0.232 0.198

Notes: Parameter estimates of bivariate MS-VECM. The regime-dependent vector $v\,(s_t)$ contains the intercept terms, the regime-dependent error-correction coefficients are given by $\alpha\,(s_t)$, and the diagonal elements (the variances) of the regime-dependent variance-covariance matrices are given by $\tilde{\Sigma}\,(s_t)$. Asymptotic standard errors in parenthesis.

Table 9: Decomposition into regime-specific attractor sets

	$\frac{R_t(6)}{R_t(3)}$	$\frac{R_t(12)}{R_t(3)}$	$\frac{R_t(36)}{R_t(3)}$	$\frac{R_t(60)}{R_t(3)}$	$\frac{R_t(120)}{R_t(3)}$
regime-dependent risk premium					
$\mu(1)$	0.157	1.010	0.672	0.658	0.694
$\mu(2)$	0.121	0.486	1.213	1.485	1.924
regime-dependent drift					
$\delta(1)$	0.045 0.045	−0.033 −0.033	0.005 0.005	0.008 0.008	0.015 0.015
$\delta(2)$	−0.020 −0.020	−0.031 −0.031	−0.014 −0.014	−0.013 −0.013	−0.013 −0.013
regime-dependent change of common trend					
$\alpha'_{\perp}(1)\delta(1)$	-0.011	0.007	-0.001	-0.001	-0.002
$\alpha'_{\perp}(2)\delta(2)$	0.003	0.003	0.001	0.001	0.000

Notes: The decomposition is derived in the appendix. The regime-dependent equilibrium mean is given by $E\left(\beta' x_t|s_t\right) = \mu(s_t) = \left[(\alpha'\alpha)^{-1}\alpha'(\Gamma C - I)v|s_t\right]$, the regime-dependent vector of drifts is given by $E(\Delta x_t|s_t) = \delta(s_t) = \left[\beta_{\perp}\left(\alpha'_{\perp}\Gamma\beta_{\perp}\right)^{-1}\alpha_{\perp}v|s_t\right]$. The expected changes of the common stochastic trend are calculated using the common factor representation of Gonzalo and Granger (1995).

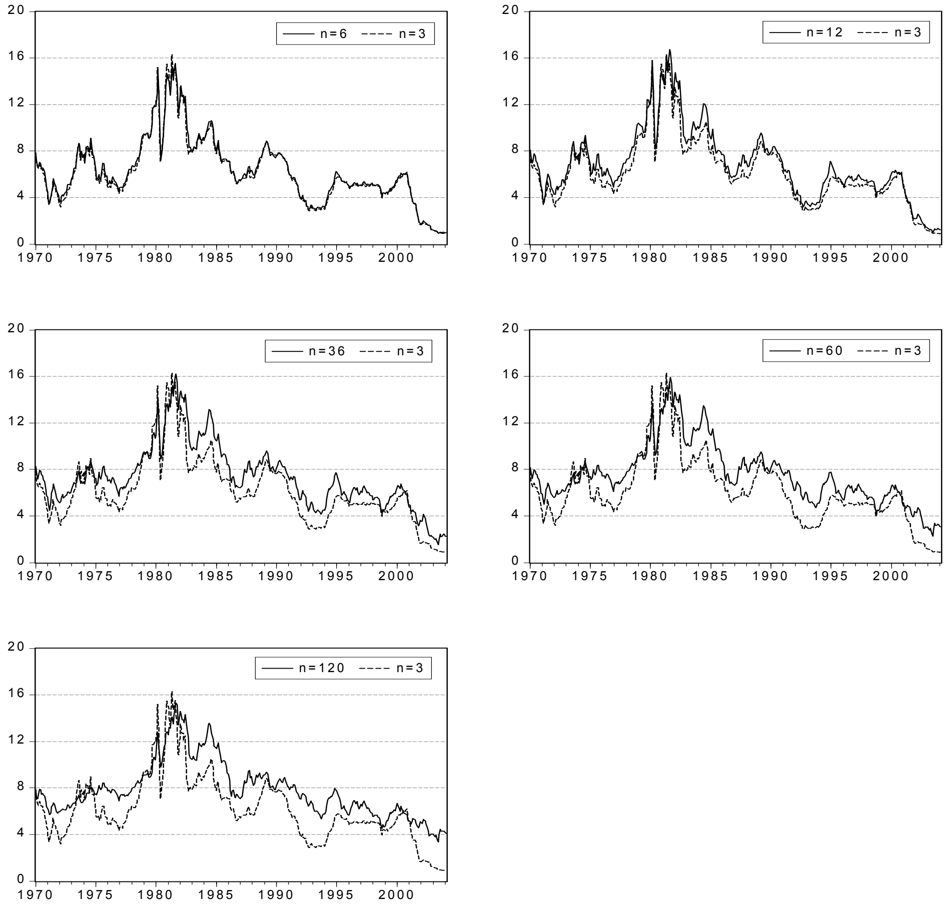

Figure 1: Interest rates (in % p.a.) on U.S. bonds of maturity n (in months), source: Federal Reserve Bank of St. Louis

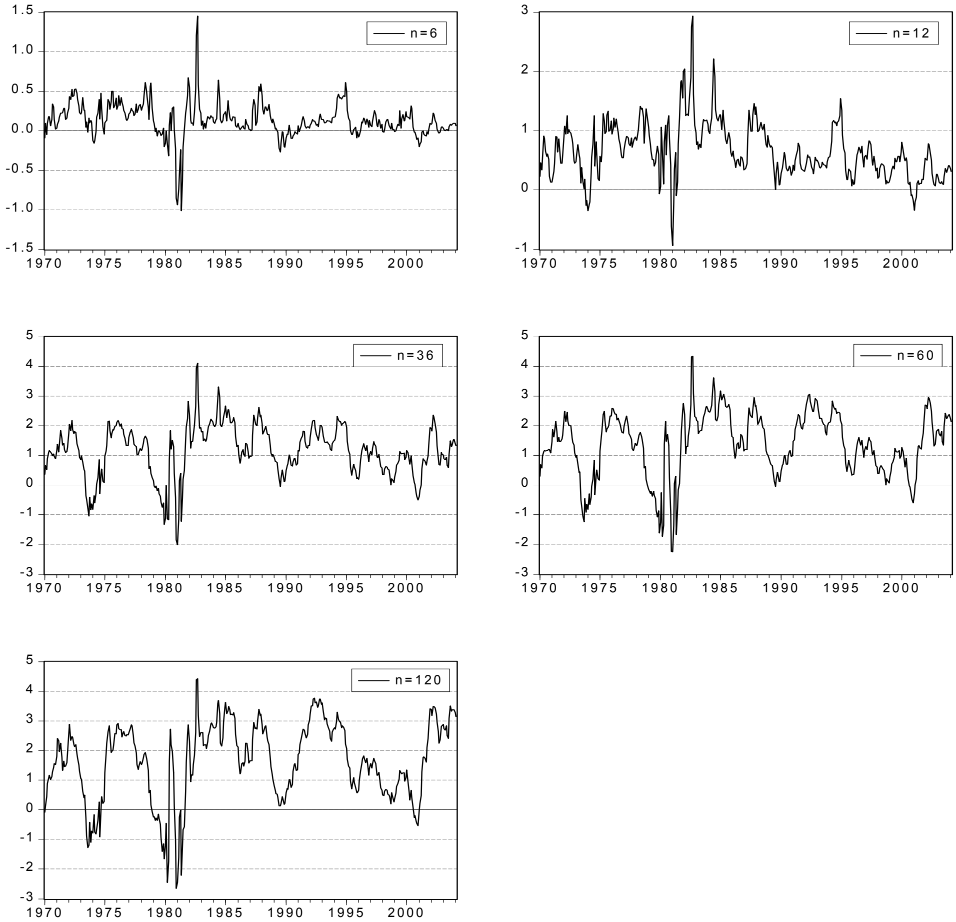

Figure 2: Spread (in percentage points) between annual interest rate on bond of maturity n (in months), $R_t(n)$, and three-months bond, $R_t(3)$

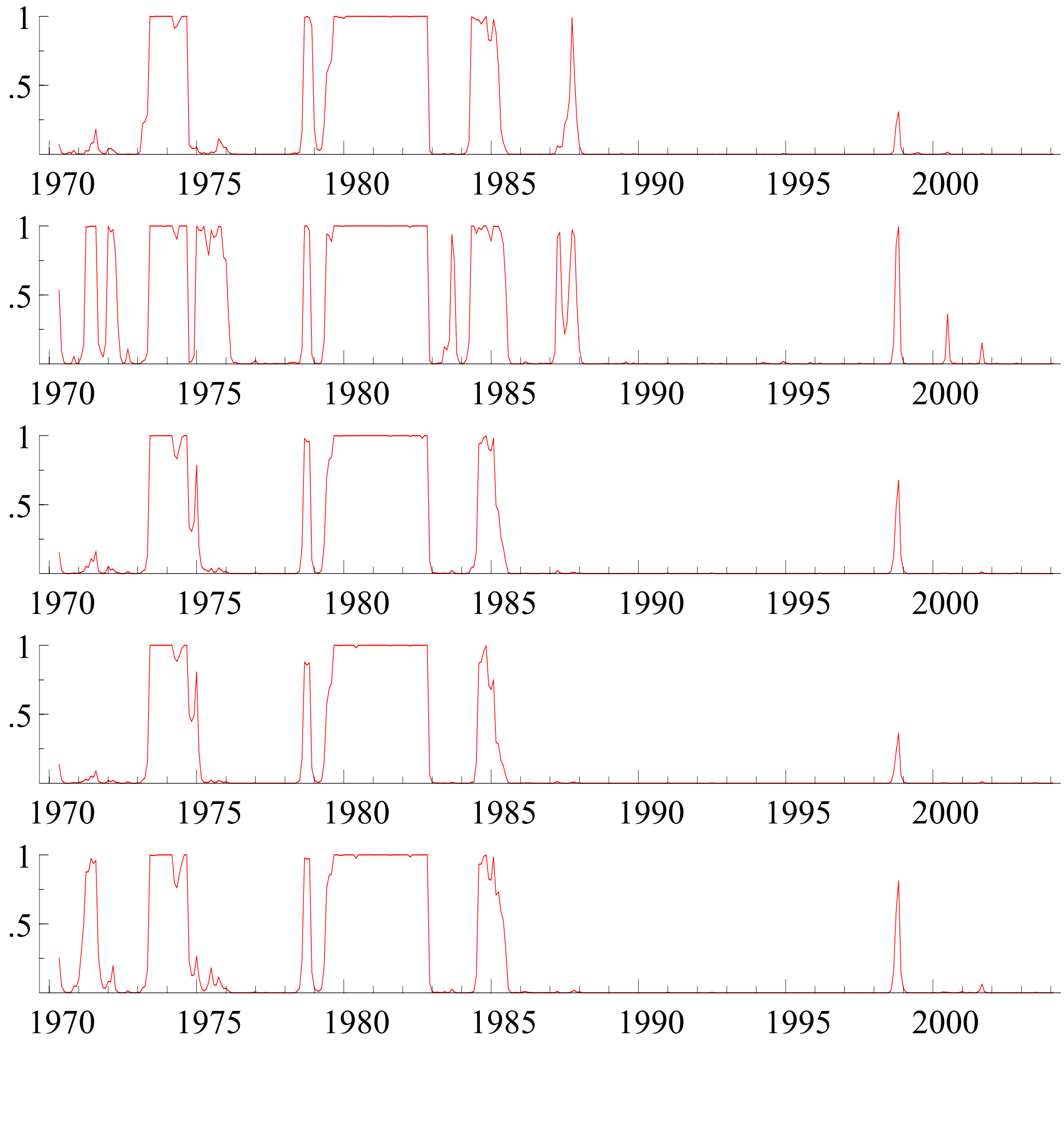

Figure 3: Conditional (smoothed) probability of regime 1 (high variance) obtained from bivariate MS-VECM for long rate of maturity $n \in \{6, 12, 36, 60, 120\}$ (from top to bottom) and short rate ($n = 3$)