Bonn Econ Discussion Papers

Discussion Paper 27/2003

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

by

Peter Tillmann

December 2003

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

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: December 2003

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. Moreover, the regime shifts in the risk premium and the equilibrium adjustment reflect shifts in monetary policy.

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

JEL classification: E43, E52

[1] I thank Katrin Assenmacher-Wesche, Jens Clausen, and Manfred J.M. Neumann as well as participants of 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 instead focuses on the long-run implications of the expectations hypothesis, that is, the necessary conditions for the expectations hypothesis to hold. The expectations 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 term structure 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 models are not capable to shed light on shifting risk premia and other regime-dependent dynamics which are likely to be induces by regime 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.

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 a regime with a high risk premium (that grows with maturity of the long bond) and a strong drift that prevails during the non-borrowed reserve-targeting episode of Federal Reserve policy during 1979-1982. Furthermore, the adjustment of long rates towards the equilibrium yield spread is much faster when the term premium and the interest rate volatility are high. Thus we supplement recent findings of e.g. Hansen (2003) who identifies shifts in risk premia and short-run dynamics 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 two strands of the literature, namely on the cointegrating properties and the regime-shifting behavior. 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 in light of the theory. 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 $F_t(1) = R_t(1)$. The Fisher-Hicks formula gives

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

The expectations hypothesis says that

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

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

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

with $\phi\left(n\right) = \frac{1}{n}\sum_{j=0}^{n-1} E_t\theta\left(j\right)$ as the expected average term premium. The pure expectations hypothesis requires $\phi\left(n\right) = 0$, while weaker versions restrict this term to be constant. 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

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

where Δ is the difference operator. From this expression we can derive the cointegration properties. 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 (provided 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 $\phi\left(n\right)$ 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 $R_t(n) - R_t(3)$ for $n \in \{6, 12, 24, 60, 120\}$ 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) who estimate a VECM system for twelve variables and find a cointegration vector consistent with the theory. Shea (1992) examines pairwise cointegration relations and finds mixed evidence for very wide horizons. Engsted and Tanggaard (1994) also 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 regime 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 is likely to have induced a change in the stochastic process of the entire term structure.
Sola and Driffill (1994), among others, 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] Moreover, they cannot reject the short-run rational-expectations restrictions implied by the expectations hypothesis. Similar studies by Kugler (1996) and Engsted and Nyholm

[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.

(2000), among many others, for Swiss and Danish data support the regime-dependent behavior of the term structure but provide mixed results on the validity of the expectations hypothesis.

Thus, 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, they disregard the long-run equilibrium relations prevailing in the levels of the series and implied by the expectations hypothesis. Thus far the cointegrating properties and the regime shifts are treated separately.[4] This paper, on the contrary, 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 over the period 1960 to 1990. 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 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 (and the cointegration rank) 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 and dates their shifts without imposing a priori break dates.

Related studies argue that the term structure is characterized by non-linear and

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

asymmetric adjustment towards the equilibrium in the sense that a regime-shift occurs once the spread crosses a certain threshold. Hansen and Seo (2002) and Seo (2003) develop a threshold cointegration model and find evidence of non-linear mean reversion. Their approach requires setting thresholds prior to estimation.[5] Enders and Siklos (2001) examine a similar question within a residual-based test for cointegration. 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 rest of this paper.

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 the widely used data set constructed by McCulloch and Kwon (1993) from raw data on U.S. Treasury bills and extended by Gregory Duffee to include

[5] Sarno and Thornton (2003) analyze the asymmetric and non-linear relationship between the federal funds rate and the three-months Treasury bill rate within a threshold error-correction model.

data through 1998.[6] The monthly data set covers the period 1970:01 to 1998:12, giving a total of 348 data points, and comprises annual yields on pure discount (zero coupon) bonds for maturity (in months) $n \in \{3, 6, 12, 24, 60, 120\}$. In terms of quality and consistency this data set is unique and closely matches the requirements of the expectations theory. A plot of each series is presented in figure (1).

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. Moreover, the Kwiatkowski-Phillips-Schmidt-Shin test rejects the hypothesis of stationarity at high significance levels. The yield spreads $R_t(n) - R_t(3)$, in contrast, are $I(0)$ since the tests reject the hypothesis of a unit root at the highest level of significance. Thus, the spread is a stationary linear combination of yields as required by the expectations hypothesis. The interest rate spreads are depicted in figure (2).

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 (1988, 1991) maximum likelihood procedure for a linear VECM. In a subsequent section the model is extended to include regime-dependent coefficients given the cointegrating properties found in the first step.

Assume that we can describe the pairwise dynamics of long- and short-term interest rates by a bivariate VAR(1) system

$$x_t = v + \Gamma_1 x_{t-1} + \varepsilon_t \tag{6}$$

with $x_t = (R_t(n), R_t(3))'$ and normally distributed Gaussian innovations $\varepsilon_t \sim IID\ N(0, \Sigma)$. 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) with $\Pi = \Gamma_1 - I_n$ and the difference operator Δ

$$\Delta x_t = v + \Pi x_{t-1} + \varepsilon_t. \tag{7}$$

[6] This data set is publicly available under `http://faculty.haas.berkeley.edu/duffee/affine.htm`.

Given that the variables in x_t are $I(1)$ Johansen (1988, 1991) formulates the hypothesis of cointegration as a reduced rank of the Π matrix

$$H(r) : rank\,(\Pi) \leq r \tag{8}$$

with

$$\Pi = \alpha\beta', \tag{9}$$

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 while the adjustment towards the equilibrium is driven by the vector of loadings α.

Estimating the cointegrated VAR model requires setting a lag order q. The standard Akaike and Schwartz criteria (AIC and SC) reported in table (1) recommend different lag orders (the AIC suggests $q = 2$ while the SC suggests $q = 1$). Both tests compare the goodness of the fit of maximum likelihood estimations and correct for the loss of degrees of freedom when additional lags are added. Since subsequent models will be heavily parameterized we favour a parsimonious specification. For this reason we follow the SC and set $q = 1$ in the VAR system because the SC uses the higher penalty for extra coefficients. 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 in the data. Second, the restricted constant in the cointegrating space closely corresponds to the risk premium derived in the theoretical discussion presented above.[7]

The results of Johansen's (1988) maximum likelihood estimation of the $\Pi = \alpha\beta'$ matrix are presented in table (3). 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 between all yield pairs and can safely set $r = 1$ in subsequent estimations.

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

To test the implications of the expectations hypothesis we impose restrictions onto the cointegrating vector $\beta = (\beta_{long}, \beta_{short})'$. In table (4) we normalize $\beta_{long} = 1$ and impose the restriction $\beta' = (1, -1)$ on the system. This restriction cannot be rejected in all cases using Likelihood Ratio tests. 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. Interestingly, the constant in the cointegrating space grows with maturity. While at the short end of the term structure the long-run equilibrium is given by $R_t(6) - R_t(3) = 0.18$, the constant grows to 1.69 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. Thus, the term 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 clearly inconsistent with the short-run implications of the expectations hypothesis but is totally in line with the existing empirical evidence. Campbell and Shiller (1991) and Hardouvelis (1994), among others, find that changes in short rates are positively correlated with the spread while longer yields react negatively to a widening spread in the short run.

Testing for weak exogeneity of either interest rate series amounts to restricting the respective adjustment coefficient to be equal to zero. The hypothesis $\alpha_{long} = 0$ can be rejected for all maturities while the hypothesis $\alpha_{short} = 0$ cannot be rejected (although the significance of rejecting increases with the maturity of the long rate). Thus, the short rate appears to be weakly exogenous while the long rate adjusts towards the equilibrium. This is not surprising since the short rate can be interpreted as the instrument of monetary policy set autonomously.[8]

Before estimating the regime-switching model we test whether the residuals from

[8] Sarno and Thornton (2003) employ the federal funds rate as the short rate and find that most of the equilibrium adjustment occurs through movements in the federal funds rate and not the long rate.

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 tests 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 the 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 (1988) models long-term properties for linear systems. However, recent work by Saikkonen (1992) and 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 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), and Clarida et al. (2003) and proceed in two steps by imposing the cointegrating properties derived in the linear model onto the regime-switching model.

[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.

[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.

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 $m \in \{1, 2\}$

$$\begin{aligned} \Delta x_t &= v(s_t) + \Pi x_{t-1} + \sum_{i=1}^{q-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \\ \varepsilon_t &\sim IID\ N(0, \Sigma(s_t)). \end{aligned} \tag{10}$$

In contrast to the model in (7), the vector of intercept terms $v(s_t)$ and the variance-covariance terms $\Sigma(s_t)$ of the innovations of this VECM are conditioned on the realization of the state variable.

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

$$\Delta x_t - \delta(s_t) = \alpha\left[\beta' x_{t-1} - \mu(s_t)\right] + \sum_{i=1}^{q-1} \Gamma_i \left[\Delta x_{t-i} - \delta(s_t)\right] + u_t \tag{11}$$

with

$$\begin{aligned} \delta(s_t) &= \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp v(s_t) = E(\Delta x_t) \\ \mu(s_t) &= -(\beta'\alpha)^{-1}[\beta' v(s_t)] = E\left(\beta' x_{t-1}\right), \end{aligned} \tag{12}$$

where 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 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 refer to this specification as MSIH-VECM.

In an alternative specification which we will refer to as MSIAH-VECM we also let the vector of adjustment coefficients α to be regime-dependent. Thus, we relax the assumption of linear adjustment towards the equilibrium

$$\Delta x_t - \delta(s_t) = \alpha(s_t)\left[\beta' x_{t-1} - \mu(s_t)\right] + \sum_{i=1}^{q-1} \Gamma_i(s_t)\left[\Delta x_{t-i} - \delta(s_t)\right] + u_t. \tag{13}$$

The decomposition into attractor sets now includes $\alpha(s_t)$ instead of α. Hamilton (1988) proposes the application of unobservable Markovian chains as regime-generating processes

$$prob(s_t = j|s_{t-1} = i, s_{t-2} = k, ...) = prob(s_t = j|s_{t-1} = i) = p_{ij}. \qquad (14)$$

In the class of models applied here the regime that prevails at time t is unobservable. The Markov property described in equation (14) 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

$$P = \begin{bmatrix} p_{11} & p_{21} = 1 - p_{22} \\ p_{12} = 1 - p_{11} & p_{22} \end{bmatrix}, \qquad (15)$$

where the element of the i-th row and the j-th column 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 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

$$L(\theta|Y_T) = \sum_{t=1}^{T} \ln f(y_t|Y_{t-1};\theta) \qquad (16)$$

with

$$\begin{aligned} f(y_t|Y_{t-1};\theta) &= f(y_t, s_t = 1|Y_{t-1};\theta) + f(y_t, s_t = 2|Y_{t-1};\theta) \qquad (17) \\ &= \sum_{m=1}^{2} f(y_t|s_t = m, Y_{t-1};\theta) \cdot prob(s_t = m|Y_{t-1};\theta), \end{aligned}$$

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

$$\widehat{\theta}_{ML} = \arg\max L(\theta|Y_T), \qquad (18)$$

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

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

4.2 Results

The parameters of the Markov chain and some diagnostic tests of both VECM specifications are given in table (7). First, 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 the 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. To test the quality of the nonlinear model against the corresponding linear VECM, Likelihood Ratio (LR) tests are usually applied that are asymptotically $\chi^2(r)$ -distributed with r degrees of freedom. However, the LR test under normal conditions does not apply here due to the existence of unidentified nuisance parameters under the alternative (the transition probabilities are not identified under the linear model). The use of the standard χ^2 distribution would therefore cause a bias of the test against the null.[12] To circumvent the problem of unidentified nuisance parameters, a cautious approach is used in this study. This implies that the LR test statistic is compared to a $\chi^2(r+n)$ distribution where n stands for the number of nuisance parameters.[13] Since the test statistic exceeds the critical value even under this conservative benchmark, the null-hypothesis can be rejected at high significance levels. In addition, the test proposed by Davies (1977) confirms the model specification.[14] Although these test statistics must be interpreted somewhat cautiously, a non-linear regime switching specification seems to be not only appropriate but rather superior to conventional linear models.[15] Allowing for shifts in the adjustment vector provides only a marginally better fit of the model. Therefore, we treat both the MSIH- and the MSIAH-VECM specifications as complimentary approaches.

The model endogenously separates distinct regimes characterized by a regime-specific vector of intercepts and a regime-specific variance-covariance matrix of the residuals.[16] The characteristics of each regime are presented in tables (8)

[12] See Hansen (1992b), Andrews and Ploberger (1994), and Garcia (1998) for this problem.

[13] The LR test statistic is computed as $LR = 2(\ln L(\theta|Y_T) - \ln L(\theta^{restr}|Y_T))$ where θ^{restr} denotes the set of parameters obtained from an estimation of the restricted (linear) VECM model.

[14] Note that the p-values for this test must be interpreted with caution since the conditions under which the test works are violated here.

[15] Likelihood Ratio tests also reject the restriction of a regime-invariant variance-covariance matrix. Similarly, estimations with more than two regimes do not generally improve the fit of the model.

[16] From the estimated transition probabilities we can derive the expected duration of each unobservable regime. While regime 2 exhibits a mean duration of about 30 months, regime 1

and (9). The vector of adjustment parameters in the MSIH-VECM is roughly similar to the linear case and the aforementioned discussion of the sign of the adjustment equally applies here. The estimated adjustment coefficients for the first two yield pairs are all positive while the long rate adjusts negatively in the linear VECM. This is clearly an improvement due to the acknowledgement of regime-dependent dynamics. However, only a few coefficients are significantly different from zero.

Allowing for shifts in the α vector in the MSIAH specification improves the results and provides interesting insights into the dynamics of the term structure in periods of high volatility and high risk premia. For all pairs the adjustment of the long rate is significant (although negative) in regime 1. Here the adjustment of the long-rate is generally stronger than in regime 2. Thus, interest rates adjust much faster towards the equilibrium in periods of unusual volatility which, as we will see in a moment, correspond to the 1979-82 period of money targeting: $|\alpha_{long}(1)| > |\alpha_{long}(2)|$. The short-rate, on the contrary, exhibits a slower adjustment which is consistent with the weak exogeneity found before. We can conclude that in regime 1 the yield spread in period $t-1$ contains more information for the course of the long rate in period t than in regime 2. As mentioned before, these results shed some light on non-linear behavior over time while previous studies analyzed asymmetric equilibrium adjustment related to the sign and the size of deviations.

The regime-dependent variance-covariance matrices exhibit a clear tendency for the variance of both the long and the short rate to decrease as the horizon of the term spread increases. Regime 1 is characterized by a much higher variance of both the long and the short rate compared to regime 2. Thus, shifts in the underlying regime substantially affect the volatility of interest rates.[17]

The regime-dependent vector of intercept terms can be decomposed into regime-specific attractor sets as explained above. These equilibrium means μ and drift terms δ are presented in table (10). As the preceding discussion made clear, the constant in the cointegrating space corresponds to the risk premium embedded in the long rate. Disentangling the regime-dependent constant of the VECM into a regime dependent mean of the equilibrium relation and a vector of drifts thus results in a regime-dependent risk premium given by $\mu(s_t)$ with extremely interesting properties. First, the risk premium in the MSIH-VECM grows with

is considerably less persistent. The results are very similar across maturities (as can be gauged from the conditional regime probabilities) and are therefore not reported here.

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

maturity in each regime from $\mu = 0.40$ for regime 1 at the short end to $\mu = 3.50$ for the widest horizon and $\mu = 0.20$ for regime 2 at the short end to $\mu = 0.98$ for the widest horizon. This is consistent with the risk premia generated by the linear VECM ranging between 0.18 and 1.69. Second, in regime 1 the long rate always entails a premium over the short rate that is always higher than in regime 2: $\mu(1) > \mu(2)$. The difference between μ across regimes grows with maturity. As stressed by Hansen (2003), the shifts in monetary policy have an important impact on the stochastic properties of interest rates and lead to substantial variation in term premia. In the MSIAH specification, risk premia are generally lower than in the case of the MSIH-VECM. However, the same tendency for the premia to increase with maturity in both regime is present, although not as pronounced as in the first case. In general, the term premium rises in times of rising volatility.[18] 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 poor 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.

Regime 1 exhibits a stronger short-term (negative) drift than regime 2 in both specifications: $E[\Delta x_t|1] < E[\Delta x_t|2]$. Hence, the adjustment of interest rates is much stronger during periods of high volatility and high risk premia.

Figure (3) presents the conditional (smoothed) regime probabilities resulting from the estimation of the MSIH model. The probabilities obtained from the MSIAH specification are virtually identical and are not reported here in order to save space. For all maturities the probabilities exhibit a strong degree of co-movement and indicate very similar regime shifts. The most important 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. However, the virtue of the regime-switching method is the ability to let the model detect regime shifts endogenously. The regime shifts in the term structure occurring in 1973/74 and 1984 are also found, among others, by Ang and Bekaert (2002). The shift to regime 1 in 1973 is more pronounced at the short end of the term structure and reflects tensions in US bond markets during the final breakup of the Bretton Woods system of fixed exchange rates. Note that the remarkable stability of regime 2 in the last third of the sample

[18] The evolution of term premia is generally consistent with the findings of Hansen (2003). Premia are higher during the 1979-1982 period than before and decrease afterwards.

that coincides with the chairmanship of Alan Greenspan since 1987.
According to the Fisher equation, yields on long term bonds reflect long run inflationary expectations given a constant real interest rate. Moreover, inflationary expectations are a major determinant of term premia. Thus, shifts to regime 1 correspond to shifts towards a regime of higher inflationary expectations. The dates of shifts to regime 1 correspond to the narrative account of recent US monetary policy of Goodfriend (1993, 1998). He identifies periods of "inflation scare" accompanied by sharply rising long rates and decreasing anti-inflation credibility. Following periods of persistent inflationary expectations during the early 1970s, the Federal Reserve engaged in aggressive disinflationary policy. However, according to Goodfriend (1998), inflationary expectations rose again in 1984. This re-emergence of inflation scare is reflected in the shift towards regime 1, see figure (3). Throughout the last decade, regime 2 prevails indicating persistent anti-inflation credibility and a stable monetary environment.
To summarize, we find that the term structure is subject to structural shifts induced by monetary policy. A shift to regime 1 raises the term premium, increases volatility, and strengthens the adjustment of long rates towards the equilibrium yield spread. Regime 1 prevails during the 1979-1982 episode and other periods of rising inflationary expectations.

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 paper argues that we can gain additional insights about the behavior of interest rates by studying these two issues jointly. In particular, the regime-switching dynamics of stationary term premia and the time-varying nature of equilibrium adjustment 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 a regime prevails that exhibits a high risk premium, a strong short-run drift, and a much faster equilibrium adjustment than in previous and subsequent episodes. 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(1) in error-correction form where we drop the regime-dependence for convenience

$$\Delta x_t = v + \alpha\beta' x_{t-1} + \varepsilon_t,$$

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, so that one of them belongs to the cointegration space determined by α and the other to Δx_t. Use Johansen's (1995, p. 39) "beautiful relation"

$$I = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp + \alpha \left(\beta'\alpha\right)^{-1} \beta',$$

where I is the identity matrix and the orthogonal complements $\alpha_\perp$ and $\beta_\perp$ have full rank r and are given by $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. Multiplying this expression with the vector v leads to

$$v = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp v + \alpha \left(\beta'\alpha\right)^{-1} \beta' v.$$

Substituting this expression into the VECM model

$$\begin{aligned}
\Delta x_t &= \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp v + \alpha \left(\beta'\alpha\right)^{-1} \beta' v + \alpha\beta' x_{t-1} + \varepsilon_t \\
&= \alpha \begin{bmatrix} \beta' & \left(\beta'\alpha\right)^{-1} \beta' v \end{bmatrix} \begin{bmatrix} x_{t-1} \\ 1 \end{bmatrix} + \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp v + \varepsilon_t
\end{aligned}$$

and summarizing yields

$$\begin{aligned}
\Delta x_t - \delta &= \alpha \left[\beta' x_{t-1} - \mu\right] + \varepsilon_t \\
\text{with } \delta &= \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp v \\
\mu &= -\left(\beta'\alpha\right)^{-1} \beta' v.
\end{aligned}$$

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

$$\begin{aligned}
E\left(\Delta x_t\right) &= \delta = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp v \\
E\left(\beta' x_{t-1}\right) &= \mu = \left(\beta'\alpha\right)^{-1} \beta' v.
\end{aligned}$$

In the MS-VECM context this procedure 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.

References

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

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

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

[4] 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.

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

[6] 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.

[7] 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.

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

[9] Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika* 64, 247-254.

[10] Enders, W. and P. L. Siklos (2001): "Cointegration and Threshold Adjustment", *Journal of Business and Economic Statistics* 19, 166-176.

[11] 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.

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

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

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

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

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

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

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

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

[20] 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.

[21] 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.

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

[23] 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.

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

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

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

[27] Hardouvelis, G. A. (1994): "The term structure spread and future changes in long and short rates in the G7 countries", *Journal of Monetary Economics* 33, 255-283.

[28] Johansen, S. (1988): "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control* 12, 231-254.

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

[30] Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Model*s, Oxford: Oxford University Press.

[31] 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.

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

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

[34] 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.

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

[36] McCulloch, J. H. and H.-C. Kwon (1993): "U.S. Term Structure Data, 1947-1991", *Working Paper*, No. 93-6, Ohio State University.

[37] 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.

[38] 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.

[39] Saikkonen, P. (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation", *Econometric Theory* 8, 1-27.

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

[41] Sarno, L. and D. L. Thornton (2003): "The dynamic relationship between the federal funds rate and the Treasury bill rate: An empirical investigation", *Journal of Banking and Finance* 27, 1079-1110.

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

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

[44] 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.

[45] 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: Choosing the lag order of the VAR system

		AIC(q)	SC(q)
$x_t' = [R_t(6), R_t(3)]$	$q = 1$	1.32	1.38*
	$q = 2$	1.31	1.42
	$q = 3$	1.33	1.48
	$q = 4$	1.35	1.55
$x_t' = [R_t(12), R_t(3)]$	$q = 1$	2.29	2.36*
	$q = 2$	2.28	2.39
	$q = 3$	2.30	2.45
	$q = 4$	2.32	2.52
$x_t' = [R_t(24), R_t(3)]$	$q = 1$	2.56	2.63*
	$q = 2$	2.52	2.63
	$q = 3$	2.53	2.68
	$q = 4$	2.55	2.75
$x_t' = [R_t(60), R_t(3)]$	$q = 1$	2.57	2.64*
	$q = 2$	2.55	2.66
	$q = 3$	2.56	2.72
	$q = 4$	2.58	2.78
$x_t' = [R_t(120), R_t(3)]$	$q = 1$	2.36	2.43*
	$q = 2$	2.36	2.47
	$q = 3$	2.38	2.53
	$q = 4$	2.41	2.61

Notes: AIC(q) and SC(q) denote the Akaike information criterion and the Schwartz criterion, respectively. The lag order q is chosen (indicated by *) in order to minimize the SC.

Table 2: Unit root tests

series		ADF	PP	KPSS
$R_t(3)$	const.	-2.15	-2.24	0.46**
	no const.	-1.07	-1.04	
$R_t(6)$	const.	-2.48	-2.11	0.50**
	no const.	-1.14	-1.01	
$R_t(12)$	const.	-2.44	-2.03	0.51**
	no const.	-0.97	-0.92	
$R_t(24)$	const.	-1.89	-1.81	0.51**
	no const.	-0.90	-0.87	
$R_t(60)$	const.	-1.47	1.50	0.50**
	no const.	-0.77	-0.77	
$R_t(120)$	const.	-1.18	-1.28	0.50**
	no const.	-0.65	-0.66	
$R_t(6) - R_t(3)$	const.	-7.74***	-7.72***	0.46**
	no const.	-4.24***	-4.85***	
$R_t(12) - R_t(3)$	const.	-6.40***	-6.44***	0.09
	no const.	-4.36***	-3.96***	
$R_t(24) - R_t(3)$	const.	-4.85***	-4.82***	0.12
	no const.	-3.34***	-3.02***	
$R_t(60) - R_t(3)$	const.	-3.88***	-3.75***	0.19
	no const.	-2.70***	-2.50**	
$R_t(120) - R_t(3)$	const.	-3.61***	-3.62***	0.26
	no const.	-2.44**	-2.30**	

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. A significance level of 1%, 5%, and 10% is indicated by ***, **, and *.

Table 3: Results of Johansen cointegration test

x_t'	H_0 $rank = r$	$\lambda^{\max}$	trace test statistic	trace test 5% cv	$\lambda^{\max}$ test statistic	$\lambda^{\max}$ test 5% cv
$[R_t(6), R_t(3)]$	$r = 0$	0.15	63.25	19.96	58.57	15.67
	$r \leq 1$	0.01	4.68	9.24	4.68	9.24
$[R_t(12), R_t(3)]$	$r = 0$	0.11	44.83	19.96	40.52	15.67
	$r \leq 1$	0.01	4.31	9.24	4.31	9.24
$[R_t(24), R_t(3)]$	$r = 0$	0.07	28.66	19.96	24.74	15.67
	$r \leq 1$	0.01	3.92	9.24	3.92	9.24
$[R_t(60), R_t(3)]$	$r = 0$	0.05	21.89	19.96	18.62	15.67
	$r \leq 1$	0.01	3.27	9.24	3.27	9.24
$[R_t(120), R_t(3)]$	$r = 0$	0.05	20.86	19.96	18.23	15.67
	$r \leq 1$	0.01	2.63	9.24	2.63	9.24

Notes: Johansen test for one lag (in levels). The constant is restricted to lie in the cointegrating space. $\lambda^{\max}$ denotes the maximum eigenvalue. The trace test, the $\lambda^{\max}$ test, and the critical values are explained in detail in Johansen (1995). The 5% critical values are from Osterwald-Lenum (1992), table 1.

Table 4: Identification of the cointegrating space

x_t'	$const.$	$\beta' = [1,\ \beta_{short}]$	$H_0 : \beta_{short} = -1$	
		β_{short}	LR (χ^2)	p
$[R_t(6), R_t(3)]$	0.18 (0.08)	-1.01 (0.01)	0.67	0.41
$[R_t(12), R_t(3)]$	0.54 (0.20)	-0.99 (0.03)	0.07	0.79
$[R_t(24), R_t(3)]$	1.08 (0.40)	-0.96 (0.05)	0.46	0.50
$[R_t(60), R_t(3)]$	1.46 (0.71)	-0.95 (0.10)	0.12	0.73
$[R_t(120), R_t(3)]$	1.69 (0.87)	-0.96 (0.12)	0.05	0.82

Notes: The constant is restricted to lie in the cointegrating space. The Likelihood Ratio (LR) test statistic of the hypothesis $\beta_{short} = -1$, i.e. the cointegrating vector $(1, -1)'$, is asymptotically χ^2 distributed. The marginal significance level is given by p. Standard errors in parenthesis.

Table 5: Testing for weak exogeneity

x_t'	$\alpha' = [\alpha_{long},\ \alpha_{short}]$		$H_0 : \alpha_{long} = 0$		$H_0 : \alpha_{short} = 0$	
	α_{long}	α_{short}	LR (χ^2)	p	LR (χ^2)	p
$[R_t(6), R_t(3)]$	-0.28 (0.13)	0.02 (0.14)	3.91	0.05	0.02	0.89
$[R_t(12), R_t(3)]$	-0.14 (0.07)	0.07 (0.07)	3.95	0.05	0.92	0.34
$[R_t(24), R_t(3)]$	-0.07 (0.04)	0.06 (0.05)	3.04	0.08	1.34	0.25
$[R_t(60), R_t(3)]$	-0.04 (0.02)	0.04 (0.03)	3.73	0.05	1.67	0.19
$[R_t(120), R_t(3)]$	-0.03 (0.01)	0.04 (0.02)	4.64	0.03	2.39	0.12

Notes: The Likelihood Ratio (LR) test statistic of the hypothesis of weakly exogenous long or short rates is asymptotically χ^2 distributed. The marginal significance level is given by p. Standard errors in parenthesis.

Table 6: Linearity tests on VECM residuals

residuals from	BDS test statistic			
	$w = 2$		$w = 3$	
	$\eta = 0.5\sigma$	$\eta = 1.5\sigma$	$\eta = 0.5\sigma$	$\eta = 1.5\sigma$
$x_t = \begin{bmatrix} R_t(6) \\ R_t(3) \end{bmatrix}$	0.03*** 0.04***	0.04*** 0.05***	0.05*** 0.05***	0.09*** 0.09***
$x_t = \begin{bmatrix} R_t(12) \\ R_t(3) \end{bmatrix}$	0.02*** 0.03***	0.03*** 0.04***	0.02*** 0.04***	0.06*** 0.08***
$x_t = \begin{bmatrix} R_t(24) \\ R_t(3) \end{bmatrix}$	0.01*** 0.04***	0.01*** 0.05***	0.02*** 0.06***	0.04*** 0.09***
$x_t = \begin{bmatrix} R_t(60) \\ R_t(3) \end{bmatrix}$	0.00** 0.04***	0.04*** 0.10***	0.00** 0.04***	0.04*** 0.10***
$x_t = \begin{bmatrix} R_t(120) \\ R_t(3) \end{bmatrix}$	0.01*** 0.04***	0.04*** 0.05***	0.02*** 0.10***	0.04*** 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% and 5% is indicated by *** and **. Bootstrapped p-values indicate virtually identical levels of significance and are not reported here.

Table 7: Results from MS-VECM estimations

x'_t	max. ln $L(\theta\|Y_T)$					Markov chain	
	linear VECM	MS-VECM	LR	p (χ^2)	p $(Davies)$	p_{11}	p_{22}
Results from MSIH-VECM							
$[R_t(6), R_t(3)]$	-222.94	9.59	465.07	0.00	0.00	0.88	0.96
$[R_t(12), R_t(3)]$	-391.55	-162.49	458.12	0.00	0.00	0.90	0.97
$[R_t(24), R_t(3)]$	-438.49	-238.37	400.24	0.00	0.00	0.89	0.97
$[R_t(60), R_t(3)]$	-440.02	-268.23	343.57	0.00	0.00	0.89	0.98
$[R_t(120), R_t(3)]$	-403.30	-254.93	296.76	0.00	0.00	0.87	0.97
Results from MSIAH-VECM							
$[R_t(6), R_t(3)]$	-222.94	13.80	473.50	0.00	0.00	0.88	0.96
$[R_t(12), R_t(3)]$	-391.55	-160.29	462.53	0.00	0.00	0.90	0.97
$[R_t(24), R_t(3)]$	-438.49	-236.69	403.60	0.00	0.00	0.89	0.97
$[R_t(60), R_t(3)]$	-440.02	-266.15	347.74	0.00	0.00	0.89	0.97
$[R_t(120), R_t(3)]$	-403.30	-250.81	304.98	0.00	0.00	0.85	0.96

Notes: The Likelihood Ratio (LR) test statistic with a marginal significance level $p(\chi^2)$ is adjusted to correct for unidentified nuisance as explained in the text and is computed as $LR = 2(\ln L(\theta|Y_T) - \ln L(\theta^{restr}|Y_T))$ where θ^{restr} denotes the set of parameters obtained from an estimation of the restricted (linear) VECM model. The significance level $p(Davies)$ invokes the upper bound derived by Davies (1977).

Table 8: Results from MS-VECM: intercepts and variances

x_t	$v(1)$	$v(2)$	$\tilde{\Sigma}(1)$	$\tilde{\Sigma}(2)$
Results from MSIH-VECM				
$R_t(6)$	−0.11	−0.02	1.40	0.09
$R_t(3)$	−0.22	−0.07	1.50	0.07
$R_t(12)$	−0.04	−0.00	1.29	0.13
$R_t(3)$	−0.15	−0.09	1.62	0.09
$R_t(24)$	0.00	0.02	0.93	0.13
$R_t(3)$	−0.09	−0.04	1.61	0.09
$R_t(60)$	0.00	0.03	0.53	0.11
$R_t(3)$	−0.11	0.00	1.73	0.09
$R_t(120)$	−0.02	0.05	0.30	0.07
$R_t(3)$	−0.13	0.01	1.51	0.08
Results from MSIAH-VECM				
$R_t(6)$	0.14	−0.05	1.23	0.09
$R_t(3)$	0.01	−0.10	1.33	0.07
$R_t(12)$	0.08	−0.04	1.22	0.13
$R_t(3)$	−0.06	−0.11	1.56	0.08
$R_t(24)$	0.04	−0.01	0.92	0.13
$R_t(3)$	−0.09	−0.06	1.60	0.09
$R_t(60)$	0.02	0.01	0.52	0.11
$R_t(3)$	−0.13	−0.00	1.71	0.09
$R_t(120)$	−0.00	0.02	0.28	0.08
$R_t(3)$	−0.17	0.01	1.52	0.08

Notes: The regime-dependent vector $v(s_t)$ contains the intercept terms, and the diagonal elements (the variances) of the regime-dependent variance-covariance matrices are given by $\tilde{\Sigma}(s_t)$.

Table 9: Results from MS-VECM: error-correction coefficients

x_t'	$\alpha(1)$		$\alpha(2)$	
	α_{long}	α_{short}	α_{long}	α_{short}
Results from MSIH-VECM				
$[R_t(6), R_t(3)]$	0.13 (0.11)	0.40 (0.09)		
$[R_t(12), R_t(3)]$	0.01 (0.06)	0.19 (0.05)		
$[R_t(24), R_t(3)]$	-0.03 (0.03)	0.06 (0.03)		
$[R_t(60), R_t(3)]$	-0.03 (0.02)	0.01 (0.02)		
$[R_t(120), R_t(3)]$	-0.03 (0.01)	0.01 (0.02)		
Results from MSIAH-VECM				
$[R_t(6), R_t(3)]$	-0.66 (0.09)	-0.29 (0.00)	0.25 (0.05)	0.49 (0.00)
$[R_t(12), R_t(3)]$	-0.26 (0.07)	-0.02 (0.00)	0.07 (0.08)	0.23 (0.00)
$[R_t(24), R_t(3)]$	-0.12 (0.05)	0.04 (0.00)	-0.00 (0.02)	0.07 (0.00)
$[R_t(60), R_t(3)]$	-0.08 (0.04)	0.07 (0.00)	-0.02 (0.02)	0.01 (0.00)
$[R_t(120), R_t(3)]$	-0.09 (0.03)	0.05 (0.00)	-0.01 (0.01)	0.01 (0.00)

Notes: The regime-invariant α in the MSIH-VECM specification is listed under $\alpha(1)$ for convenience. The vector of adjustment coefficients is given by $\alpha(s_t)$. Standard errors in parenthesis.

Table 10: Decomposition of VECM constant into attractor sets

x_t	$\mu(1)$	$\mu(2)$	$diff.$	$\delta(1)$	$\delta(2)$
Results from MSIH-VECM					
$\begin{bmatrix} R_t(6) \\ R_t(3) \end{bmatrix}$	0.40	0.20	0.20	-0.27 -0.27	-0.10 -0.10
$\begin{bmatrix} R_t(12) \\ R_t(3) \end{bmatrix}$	0.60	0.47	0.27	-0.03 -0.03	-0.00 -0.00
$\begin{bmatrix} R_t(24) \\ R_t(3) \end{bmatrix}$	1.07	0.72	0.35	-0.03 -0.03	-0.00 -0.00
$\begin{bmatrix} R_t(60) \\ R_t(3) \end{bmatrix}$	2.70	0.85	1.85	-0.08 -0.08	0.01 0.01
$\begin{bmatrix} R_t(120) \\ R_t(3) \end{bmatrix}$	3.50	0.98	2.52	-0.11 -0.11	0.02 0.02
Results from MSIAH-VECM					
$\begin{bmatrix} R_t(6) \\ R_t(3) \end{bmatrix}$	0.39	0.20	0.19	-0.10 -0.10	0.00 0.00
$\begin{bmatrix} R_t(12) \\ R_t(3) \end{bmatrix}$	0.56	0.50	0.06	-0.07 -0.07	0.00 0.00
$\begin{bmatrix} R_t(24) \\ R_t(3) \end{bmatrix}$	0.73	0.69	0.04	-0.20 -0.20	-0.01 -0.01
$\begin{bmatrix} R_t(60) \\ R_t(3) \end{bmatrix}$	1.00	0.64	0.36	-0.06 -0.06	0.00 0.00
$\begin{bmatrix} R_t(120) \\ R_t(3) \end{bmatrix}$	1.15	0.49	0.66	-0.10 -0.10	0.01 0.01

$Notes$: The decomposition is derived in the appendix. The regime-dependent equilibrium mean is given by $\mu(s_t) = -(\beta'\alpha)^{-1}[\beta' v\,(s_t)]$, the regime-dependent vector of drifts is given by $\delta\,(s_t) = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp v\,(s_t)$. The difference $\mu\,(s_t = 1) - \mu\,(s_t = 2)$ is denoted by $diff.$

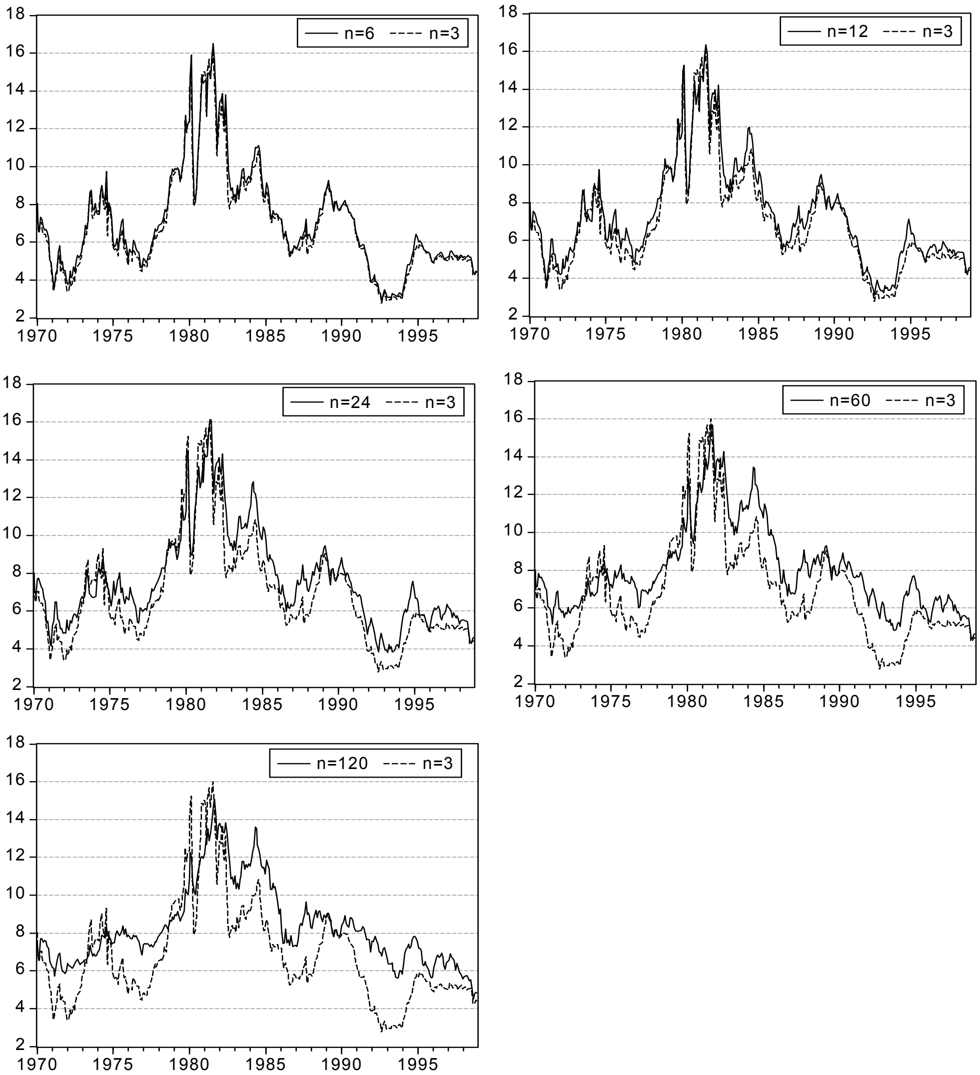

Figure 1: Yields (% p.a.) on pure discount bonds of maturity n, 1970:01-1998:12

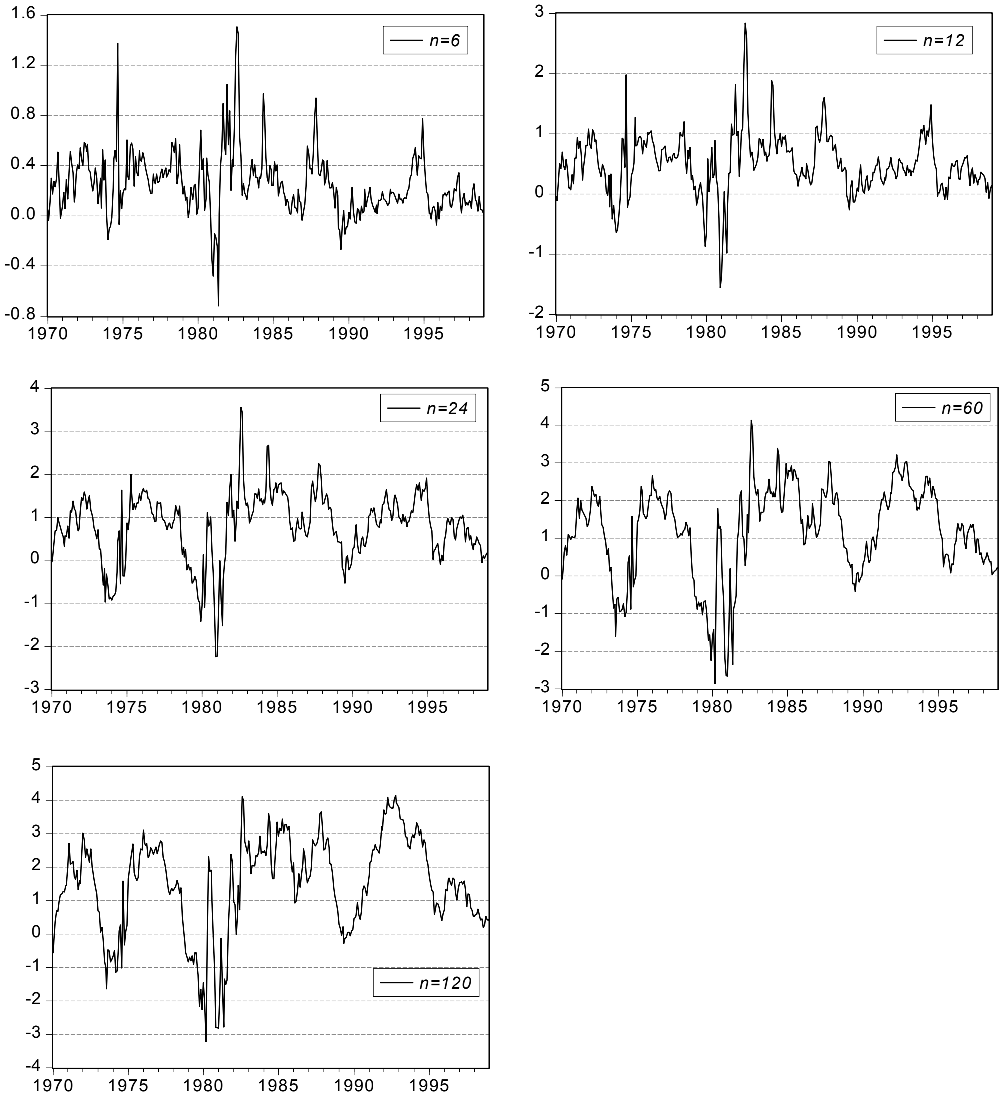

Figure 2: Spread (in percentage points) between annual yields on bond of maturity n, $R_t(n)$, and three-months bond, $R_t(3)$.

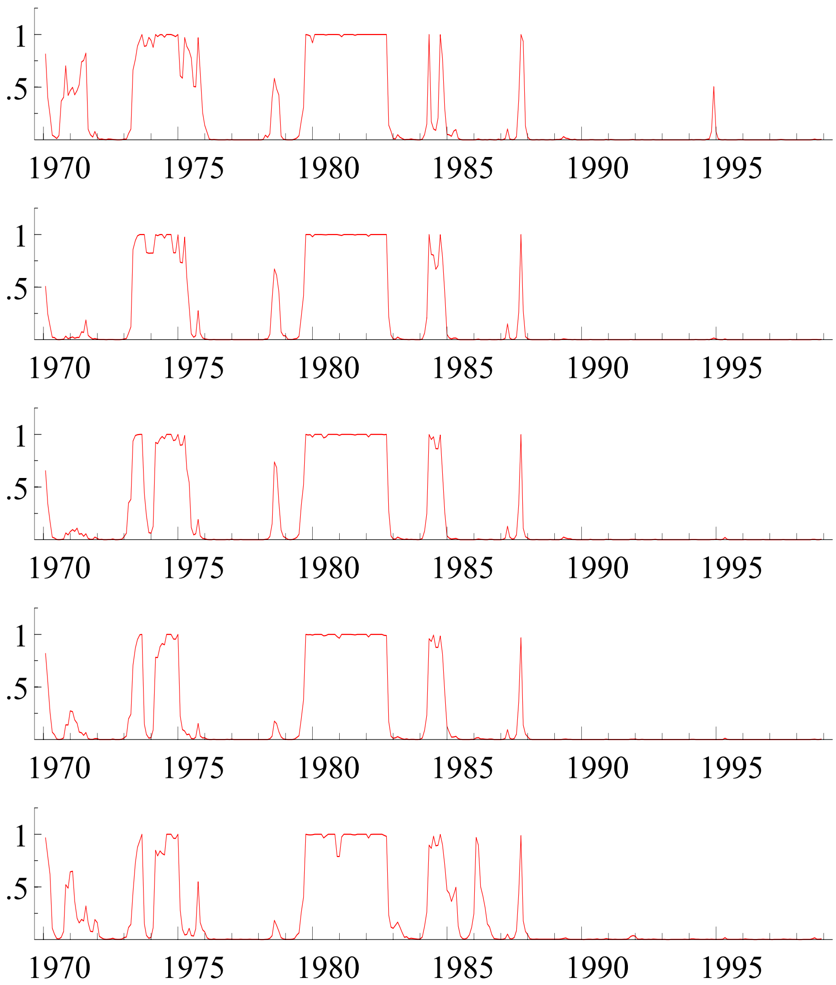

Figure 3: Conditional (smoothed) probability of regime 1 obtained from bivariate MSIH-VECM model for short rate ($n = 3$) and long rate of maturity $n = \{6, 12, 24, 60, 120\}$ months (from top to bottom)