Term Premia under Switching Regimes

by

Mårten Blix*

First version: January 1996
This version: September 1997

Abstract

In this paper we use the present value model (PVM) for the term structure of interest rates from Campbell and Shiller (1987), and extend it to include Markov switching (MS). The statistical model used is a VAR with unobservable MS from the seminal work of Hamilton, in which all the parameters in the VAR are allowed to depend on the current unobservable state. Under the null of the expectations hypothesis, the term premium conditional on the information set is time varying with the regime - and is heteroskedastic. The model is estimated on US and Swedish bond data, and we find that (1) the expectations hypothesis is still rejected; (2) the conditional term premia are large and very sensitive to the choice of discount factor; (3) the most important feature of the specification is the state dependent covariance matrix; and (4) the economic fit of the PVM is still good.

Keywords: Markov switching, expectations hypothesis, time-varying term premia.

JEL Classification numbers: C13, C32, E43.

---

* Economics Department, Sveriges Riksbank. Email: marten.blix@riksbank.se. This paper is part of research from when the author was visiting the Riksbank, and is one chapter in the author's PhD dissertation at the Institute for International Economic Studies, Stockholm University. I am indebted to Anders Warne for helpful discussions and for generously supplying his Gauss code for the EM algorithm; special thanks are also due to Tom Engsted for detailed comments and suggestions. I would like to thank Magnus Dahlquist, Hans Dillén, Helmut Lütkepohl, Torsten Persson, Lars E.O. Svensson, and Paul Söderlind for useful comments; Magnus Forsells, for help with obtaining data. I am grateful to Sveriges Riksbank for financing this research.

1 Introduction

Tests of the expectations hypothesis (EH) using linear vector autoregressions (VAR:s) have treated the term premium as a constant unexplained deviation from expectations, such as in Campbell & Shiller (1987). Most tests of the EH in the literature reject it, see the survey by Shiller (1990). In this paper we use the VAR methodology with unobservable Markov switching between discrete states, which allows a more flexible specification of interest rates and the term premium than linear VAR:s.

Markov switching models have several attractive features. First, they may fit the data better. For example, Hamilton (1988) showed that a hidden Markov model (HMM) could fit US bond data, but that a linear model could not. As discussed in Hamilton (1989), a HMM can parsimoniously represent ARIMA models with long lags. Second, sometimes the states have useful economic interpretations, either as real discrete events, or as approximations to some continuous variable.

The focus in this paper is on the term premium and its properties in a simple HMM. While a constant premium might sensibly be ignored, e.g. by central banks, a time varying premium can be potentially important. For example, a raise in the short rate may fail to have the intended effect on the yield curve if the inflation history casts doubt on the central bank's commitment to tighten monetary policy. More generally, if the level of the premium ψ_t depends on past interest rates and discrete states, then we would want to know the functional form (if any) of this dependence.

In this paper we will assume that the premium depends on the current state only, and the premium will thus be denoted by ψ_{s_t}. This is not as restrictive as it may sound, because the states are assumed to be unobserved and consequently have to be inferred from the information set. In particular, the inference of the current state will be based on past interest rates, indirectly providing a link to the monetary policy "track-record". Note further that this assumption implies that each state will have a different - but constant - premium.

The approach used is based on the present value model (PVM) of Campbell & Shiller (1987). We deviate from them by introducing an unobservable Markov Chain, that affects the drift, the autoregressive terms as well as the covariance matrix of the VAR disturbances. Applying very similar methods to Campbell & Shiller (1987), we impose restrictions on the autoregressive terms only. Under the null of the EH, the remaining term is interpreted as the premium. In other words, under null the premium is defined by our choice of statistical model.

Since the state is unobservable we will not be able to specify ψ_{s_t} directly, but we will be able to compute a premium conditional on observable data. It will depend in a simple way on uncertainty about which is the current regime, thereby generating rich time variation in the form of conditional heteroskedasticity.

HMM have been used for the term structure by Hamilton (1988), Driffill (1992), Sola & Driffill (1994), and Kugler (1996). These papers have used what might be labelled *centered* regime shift models, which differ from the one used here. The model here is much easier to estimate, has the intuitive appeal of symmetric state dependence, and is linear in parameters after conditioning on the current state. These issues are briefly discussed in Warne (1996).

These econometric advantages might be outweighed if economic theory suggested that the centered model were more appropriate, but there seems to be no particular reason why this should be the case. In terms of conditional forecasts, the two models differ only in the drift term, and economic theory typically has very little to say about such aspects of econometric models.

The rest of this paper is outlined as follows. The next section introduces the regime shift VAR model. Section 3 discusses the derivation of the conditional premium. Section 4 is the empirical part of the paper and contains a test of the PVM on US and Swedish bond data; section 5 concludes.

2 A VAR with Markov Switching

The HMM used is the same as the one in Blix (1997 a,b), which we summarise here for completeness. The model is a VAR(p) of the form

$$y_t = \mu_{s_t} + \sum_{i=1}^{p} B_{s_t}^{(i)} y_{t-i} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t | s_t \sim \mathrm{N}(0, \Omega_{s_t})$, and $s_t \in \{1,2,\cdots,q\}$ denotes the unobservable regime variable, which is assumed to follow a first order Markov Chain (MC), y_t is a $n \times 1$ vector of weakly stationary variables, $B_{s_t}^{(i)}$ is the $n \times n$ state dependent parameter matrix for the *i:th* lag, μ_{s_t} is the vector of state dependent drift terms, and Ω_{s_t} is the state dependent positive definite covariance matrix . The vector $[y_0', \cdots, y_{1-p}']'$ of initial observations is taken to be fixed in repeated sampling.

The Markov transition probabilities $p_{ij} = \text{pr}[s_t = j | s_{t-1} = i]$ are collected into

$$P = \begin{pmatrix} p_{11} & \cdots & p_{q1} \\ \vdots & & \vdots \\ p_{1q} & \cdots & p_{qq} \end{pmatrix}, \tag{2}$$

where $p_{iq} = 1 - \sum_{j=1}^{q-1} p_{ij}$, so that $1_q{}'P = 1_q{}'$, where 1_q is a column vector of ones.

We assume that all probabilities are positive, so that we have an irreducible chain. The Markov assumption implies that the only relevant information for predicting future states is the current state, so that $\text{pr}[s_t | \mathcal{Y}_{t-1}, s_{t-1,} s_{t-2}, \cdots] = \text{pr}[s_t | s_{t-1}]$, where $\mathcal{Y}_{t-1} = [y_{t-1}, y_{t-2}, \cdots]$. We further assume that the current state is not known with certainty, and collect all the probabilities of being in a particular state based on the information set $\mathcal{Y}_t$ in the $q \times 1$ vector

$$\xi_{t|t} = \begin{bmatrix} \text{pr}[s_t = 1 | \mathcal{Y}_t] \\ \vdots \\ \text{pr}[s_t = \text{q} | \mathcal{Y}_t] \end{bmatrix}. \tag{3}$$

We put the model in companion form VAR as follows

$$Y_t = J'\mu_{s_t} + B_{s_t} Y_{t-1} + J'\varepsilon_t, \tag{4}$$

where

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \qquad B_\tau = \begin{bmatrix} B_\tau^{(1)} & B_\tau^{(2)} & & B_\tau^{(p)} \\ I_n & \cdots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & I_n & 0 \end{bmatrix}, \qquad J = \begin{bmatrix} I_n & 0 & \cdots & 0 \end{bmatrix}, \tag{5}$$

which are $np \times 1$, $np \times np$, and $n \times np$ respectively. Pre-multiply (4) by J, yielding

$$y_t = \mu_{s_t} + JB_{s_t} Y_{t-1} + \varepsilon_t. \tag{6}$$

3 Term Premia and the PVM

In this section we will derive and discuss the conditional premium based on the PVM of Campbell & Shiller (1987) for a HMM. The version of the PVM we use asserts that

$$\mathrm{H}_0\colon \quad R_t = (1-\delta)\sum_{j=0}^{\infty} \delta^j \mathrm{E}\left[r_{t+j} | \mathcal{Y}_t, s_t\right] + \psi_{s_t}, \tag{7}$$

where R_t is a consol, r_t is the short rate, $\delta \in (0,1)$ is a factor of linearisation assumed to be constant. This formulation stems from the seminal work of Shiller (1979). It embodies the notion that long rates should be an average of future expected short rates. The declining weights of δ are due to the effect of coupon payments; for zero-coupon bonds a flat weighing scheme would be used instead.

The model in (7) departs from the standard PVM in a potentially important way. We have specified a null hypothesis where the inference about future short rates is made with the current regime known. The appeal of this formulation is that given the regime s_t the premium is constant, just as in standard linear models. The main feature of the specification is thus that each state is assumed to carry a different premium, and that the premium depends on the current state *only*. We can then compute $\mathrm{E}\left[\psi_{s_t} \middle| \mathcal{Y}_t\right]$ by imposing the null of the EH, but leaving the drift term unrestricted. It is possible to restrict the premium to be equal across states, such as in Sola & Driffill (1994), but this places additional restrictions on the parameters. In particular, it requires selected parameters in the drift and the AR part to be state invariant, which is a testable restriction.

The PVM that we implement is obtained by subtracting r_t from (7), which yields

$$\text{H}_0: \quad S_t = \sum_{j=1}^{\infty} \delta^j \text{E}\left[\Delta r_{t+j} \middle| \mathcal{Y}_t, s_t\right] + \psi_{s_t}, \tag{8}$$

where $S_t \equiv R_t - r_t$ is the spread between the long and the short rate. Following Shiller, Campbell, and Schoenholtz (1983), we will set $\delta = (1+\overline{R})^{-1}$, but the qualitative results (below) are not sensitive to perturbations in δ.

Note that the conditional premium under the null is undefined until we specify a statistical model for R_t and r_t. By using the VAR in section 2, we can use the estimated parameters and the state probabilities $\xi_{t|t}$ to calculate it, as will be shown below. For this purpose, let

$$y_t = \begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix}, \tag{9}$$

so that the PVM can be written as

$$\text{H}_0: \quad e_2{'} y_t = e_1{'} \sum_{j=1}^{\infty} \delta^j \text{E}\left[y_{t+j} \middle| \mathcal{Y}_t, s_t\right] + \psi_{s_t}, \tag{10}$$

where e_i is the $i{:}th$ column of an identity matrix.

The HMM in section 2 is the most general version we consider, which will be denoted by model 1. We will also explore whether or not simpler models - with less regime dependence - are warranted by the data. In particular, it is of interest to test whether the AR part (only) is state invariant (denoted by model 2). We will also test if both the drift and the AR terms are state invariant (denoted by model 3). For all three models, the covariance matrix of the disturbances will be state dependent. It would be desirable to test this aspect of the parameterisation as well, but conventional statistical tests (such as LR, LM, and Wald) include a nuisance parameter that is not identified under the null. See the references in Hamilton (1994), and the papers by Garcia (1997) and Rydén (1995) for available results.

From (10) we note that we need conditional forecasts of the vector y_{t+i}. For this purpose, substituting backwards we find that for $j \geq 2$,

$$\begin{aligned} y_{t+j} = \mu_{s_{t+j}} + J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h} B_{s_{t+j+1-m}}\right)J'\mu_{s_{t+j-h}} + J\left(\prod_{m=1}^{j} B_{s_{t+j+1-m}}\right)Y_t \\ +\varepsilon_{t+j} + J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h} B_{s_{t+j+1-m}}\right)J'\varepsilon_{t+j-h}. \end{aligned} \tag{11}$$

Using lemma 4.3 in Blix (1997, p 17), forecasts of (11) for $\tau \in \{1,\cdots,q\}$ are given by

$$\mathrm{E}\left[y_{t+j}\middle|\mathcal{Y}_t, s_t = \tau\right] = \begin{cases} aPe_\tau + bP_{np}\tilde{Y}_{t|\tau} & \text{for } j=1 \\ \left(aP^j + b\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right)e_\tau + b\Phi^{j-1}P_{np}\tilde{Y}_{t|\tau} & \text{for } j \geq 2, \end{cases} \tag{12}$$

where $\tilde{Y}_{t|\tau} = e_\tau \otimes Y_t$,

$$B = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & B_q \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mu_q \end{bmatrix}, \tag{13}$$

are $npq \times npq$ and $nq \times q$ matrices respectively, $b_\tau = \left[B_\tau^{(1)} \ \cdots \ B_\tau^{(p)}\right] = JB_\tau$,

$$\begin{aligned} a &= \left[\mu_1 \ \cdots \ \mu_q\right], & b &= \left[b_1 \ \cdots \ b_q\right], \\ \Phi &= P_{np}B, & \Psi &= \left(I_q \otimes J'\right)P_n\,\mu, \\ P_\tau &= \left(P \otimes I_\tau\right), & C_\tau &= \left(1_q' \otimes I_\tau\right). \end{aligned} \tag{14}$$

Substituting (12) into the PVM in (10), we obtain

$$\begin{aligned} e_2'JC_{np}\tilde{Y}_{t|\tau} &= e_1'\left[a\sum_{j=1}^{\infty}\left(\delta P\right)^j + b\sum_{j=2}^{\infty}\delta^j\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right]e_\tau \\ &\quad +e_1'\delta b\sum_{j=1}^{\infty}\left(\delta\Phi\right)^{j-1}P_{np}\tilde{Y}_{t|\tau} + \psi_\tau \\ &\equiv \lambda' e_\tau + b^*\tilde{Y}_{t|\tau} + \psi_\tau, \end{aligned} \tag{15}$$

since $e_2'JC_{np}\tilde{Y}_{t|\tau} = e_2'y_t$. Equating the coefficients of the $\tilde{Y}_{t|\tau}$ terms imposes the null of the EH, which defines the premium ψ_τ. As discussed in Blix (1997 a, b) the restrictions can be written

$$\mathrm{H}_0\!: \; R\,\mathrm{vec}\,b = r, \tag{16}$$

where $R = I_{npq} \otimes \delta(1 \quad 1)$, and $r = \mathrm{vec}\,e_2{}'JC_{np}$. Note that the restrictions are linear, and that they are a generalisation of Campbell & Shiller (1987) in the sense that if $q = 1$ we obtain exactly their form of term structure restrictions. The premium is then given by

$$\psi_\tau = -\lambda' e_\tau \;\text{ on }\; \mathrm{H}_0\!: \; R\mathrm{vec}\,b = r. \tag{17}$$

Thus, to compute the premium we need an estimate of λ, given in the next proposition.

Proposition 3.1

$$\begin{aligned} b^* &= e_1{}'\delta b \overline{\Phi}_1^{-1} P_{np}, \\ \lambda &= \delta E_1{}'(I_q \otimes a)\mathrm{vec}\left[P\overline{P}_1^{-1}\right] \\ &\quad + \delta^2 E_1{}'(I_q \otimes b) D^{-1} P_{npq}{}'\left(P_{npq}{}'\widetilde{P}_1^{-1} - \Phi_q \widetilde{\Phi}_1^{-1}\right)\mathrm{vec}\,\Psi, \end{aligned} \tag{18}$$

where $E_\tau = I_q \otimes e_\tau$,

$$\begin{aligned} \Phi_\tau &= I_\tau \otimes \Phi, & D &= P_{npq}{}' - \Phi_q \\ \overline{P}_\tau &= I_q - (\delta P)^\tau, & \overline{\Phi}_\tau &= I_{npq} - (\delta\Phi)^\tau \\ \widetilde{P}_\tau &= I_{npq^2} - (\delta P_{npq}{}')^\tau, & \widetilde{\Phi}_\tau &= I_{npq^2} - (\delta\Phi_q)^\tau, \end{aligned} \tag{19}$$

if $\overline{\Phi}_1$, $\overline{P}_1$, D, $\widetilde{P}_1$, and $\widetilde{\Phi}_1$ are invertible.

Proof: The derivation of b^* is straightforward; for λ it is in the appendix.

Since the state is assumed to be unobserved, we condition only on observable information $\mathcal{Y}_t$. By the law of iterated expectations, if the PVM in (10) holds for the information set given by $\{\mathcal{Y}_t, s_t\}$ it should also hold for the smaller information set $\mathcal{Y}_t$. Therefore,

$$\mathrm{E}[\psi_{s_t} | \mathcal{Y}_t] = -\lambda' \xi_{t|t} \text{ on } \mathrm{H}_0\text{: } R\,\mathrm{vec}\,b = r, \tag{20}$$

which is conditionally heteroskedastic due to the dependence on $\xi_{t|t}$.

If we instead consider model 2 in which $B_{s_t} = B$ is a $np \times np$ matrix, tests of the PVM are still given by (16) but with $R = I_{np} \otimes \delta(1 \quad 1)$ and $r = \mathrm{vec}\,e_2'J$, identical to Campbell & Shiller (1987). Also, we obtain the somewhat simpler expression

$$\lambda = \delta E_1' C\left(P_{np}' - \tilde{B}_q\right)^{-1} P_{np}'\left[P_{np}' \hat{P}_1^{-1}{}' - \tilde{B}_q \overline{B}_1^{-1}\right] C' \mathrm{vec}\,a, \tag{21}$$

where $C = I_q \otimes J$, $\tilde{B}_\tau = I_\tau \otimes B$, $\hat{P}_\tau = I_{npq} - (\delta P_{np})^\tau$ and $\overline{B}_\tau = I_{npq} - \left(\delta \tilde{B}_q\right)^\tau$. The derivation is similar to (18) and is thus omitted.

4 Empirical Implementation

For the US we choose a 30 year treasury bond obtained from the FRED database at the Federal Reserve Bank of St. Louis as the long rate R_t, which is intended to approximate a consol; and a one month Euro Market as the short rate r_t, obtained from Sveriges Riksbank. Both are sampled monthly; the short rate is available for 1963/7-1997/3, but the long rate only for 1977/2-1997/3. They are plotted for 1977/2-1997/3 in figure 1a.

Figure 1a. US Bond Yields

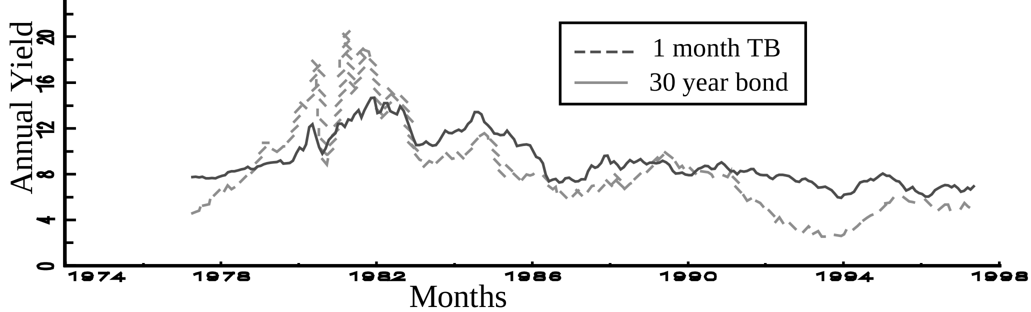
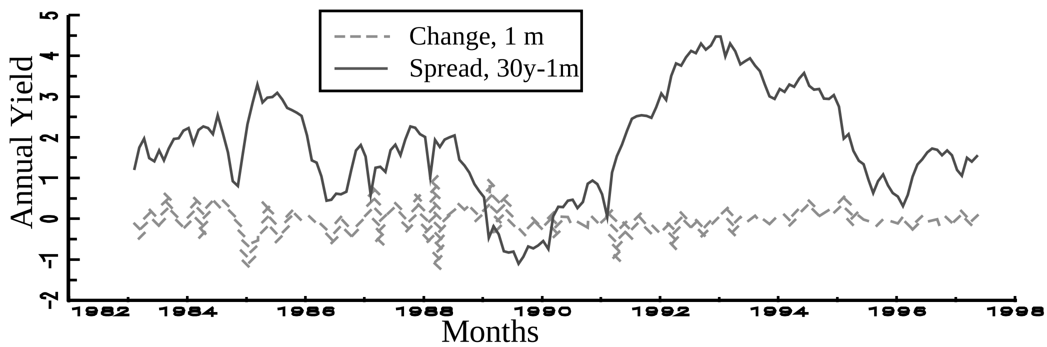


Figure 1b displays the transformed data Δr_t and $S_t \equiv R_t - r_t$ from 1982/11. Although we have data from 1977/02, we will estimate the model from 1982/11, thus omitting the Federal Reserve's monetary targeting experiment that was one of Hamilton's motivations for considering the Markov model. We do this because the Fed's experiment can be considered as a rare event that is extremely unlikely to occur again. In other words, even though we obtain a positive transition probability back to the "state" occurring 1979-82, we might sensibly suspect that the "true" transition probability is virtually zero. If the only concern is to fit the data, this might not be a problem, but since the model is used to generate forecasts to the infinite future, it would predict such an event occurring again. However, if US interest rates ever fluctuate as much as during 1979-82 it is more likely to stem from some other source not modelled (and hence unpredictable) from the VAR.

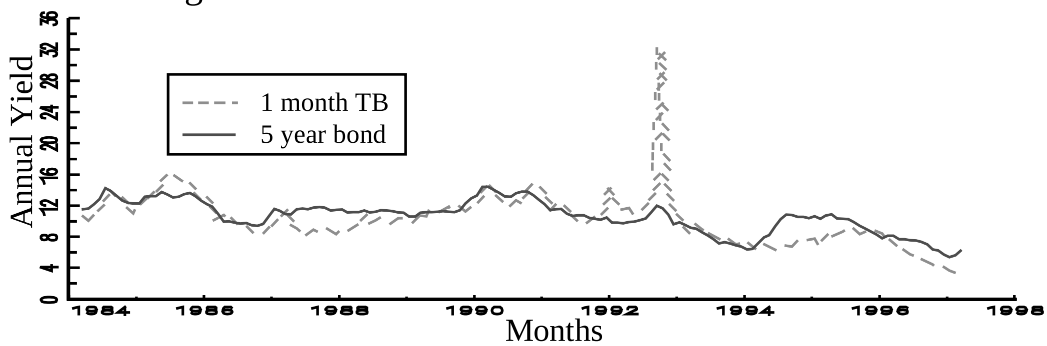
Figure 1b. US Bond Yields, Excl. Fed Experiment



Moreover, a very small transition probability may generate numerical problems, since $p_{ij} \approx 0$ is close to a corner solution in the likelihood function where a local maximum cannot be guaranteed. Finally, it might be argued that the markets knew (with probability one) that the Fed had a new operating procedure. In this case, the probability inference $\xi_{t|t}$ lies in the corner of possible outcomes, which is a problem[1] for the (EM) estimation algorithm used.

For Sweden we choose the 30 day treasury bill rate, available for 1983/01-1997/03, and the five year treasury bond rate, available for 1984/03-1997/03; both series have been obtained from Sveriges Riksbank. Five years is a rather short maturity to approximate a consol, but it is the only long bond rate available from the early eighties. Moreover, although some bond data is available from before 1984, restrictions on movements of capital make such data unsuitable for our purposes. These Swedish bond yields are plotted in figure 2a; the most notable feature of the data is the big spike in the one-month rate for September 1992. This observation occurred in the period when the Riksbank's marginal lending rate was set to 500% in an attempt to keep the exchange rate fixed. This subsequently failed and the Swedish krona was left to float.

Figure 2a. Swedish Bond Yields

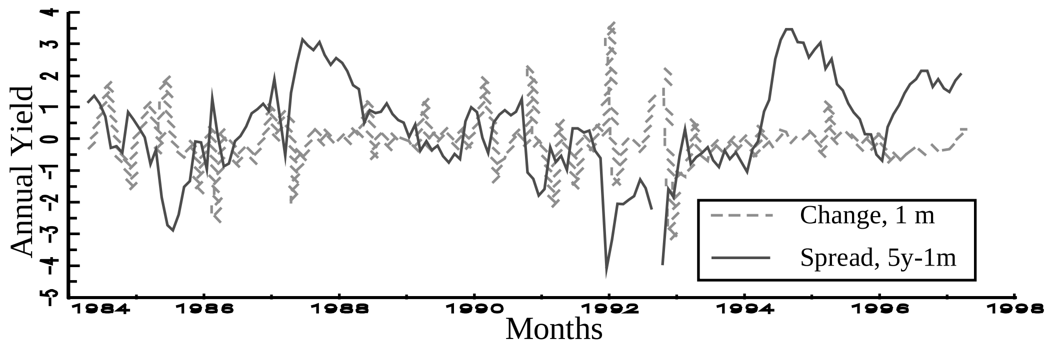


One way to view the September 1992 observation is as a "super" volatility state to which we might "jump" with some small probability. This is not an

[1] The smoothed inferences $\xi_{t|T}$ are used in the EM algorithm (see equations 4.1-4.3 in Hamilton (1990, p 51)). These smoothed probabilities are obtained from an updating formula involving division by $\xi_{t|t}$ (see equation 22.4.14 in Hamilton (1994, p 694)). If a smoothed inference is close to zero the program may crash.

approach we will pursue. We are simply going to delete this observation from the sample. This is not a satisfactory method of dealing with outliers, but appears warranted in this case. The reason for this is that it is unlikely that any trade took place when the short rate was in the 30% region. Evidence supporting this comes from the bid-ask spreads, which are usually about five basis points, but at this time reached about 1400 basis points. It therefore seems justified to omit the observation; the transformed data excluding this observation are plotted in figure 2b.

Figure 2b. Swedish Bond Yields, Excl. Sept. 1992



4.1 Estimating the unrestricted model

The model is estimated with the EM-algorithm suggested in Lindgren (1978), Hamilton (1990), and Holst, Lindgren, Holst, and Thuvesholmen (1994). We choose two regimes as the simplest possible way to capture regime switching. For all estimations we have tried a variety of starting values to confirm that the likelihood is at a local maximum.

4.1.1 Estimates for the US

For the US, we choose two lags in the VAR; a one lag VAR appeared mis-specified on criteria discussed below.

Table 1. Model 1, US 1982 / 11 - 1997 / 3.

$$L(\hat{\theta}) = 5.44, \quad \delta = 0.993, \quad \mathrm{E}[S_t] = 1.82, \quad \mathrm{E}[\Delta r_t] = -0.02, \quad P = \begin{pmatrix} \underset{(0.08)}{0.76} & \underset{(0.07)}{0.21} \\ \underset{(0.08)}{0.24} & \underset{(0.07)}{0.79} \end{pmatrix},$$

State 1

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.13)}{-0.07} \\ \underset{(0.10)}{0.10} \end{pmatrix} + \begin{pmatrix} \underset{(0.21)}{0.24} & \underset{(0.27)}{0.30} \\ \underset{(0.22)}{-0.11} & \underset{(0.26)}{0.83} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.22)}{0.02} & \underset{(0.27)}{-0.29} \\ \underset{(0.24)}{-0.20} & \underset{(0.26)}{0.13} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix},$$

$$\Omega_1 = \begin{pmatrix} \underset{(0.037)}{0.198} & \underset{(0.029)}{-0.150} \\ \underset{(0.029)}{-0.150} & \underset{(0.026)}{0.148} \end{pmatrix}$$

State 2

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.03)}{-0.04} \\ \underset{(0.06)}{0.02} \end{pmatrix} + \begin{pmatrix} \underset{(0.07)}{0.57} & \underset{(0.05)}{0.36} \\ \underset{(0.16)}{0.01} & \underset{(0.12)}{1.26} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.07)}{0.08} & \underset{(0.05)}{-0.33} \\ \underset{(0.13)}{-0.12} & \underset{(0.12)}{-0.29} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix}$$

$$\Omega_2 = \begin{pmatrix} \underset{(0.004)}{0.012} & \underset{(0.006)}{-0.006} \\ \underset{(0.006)}{-0.006} & \underset{(0.041)}{0.057} \end{pmatrix}.$$

Misspecification Tests

	Equation 1	Equation 2
Autocorrelation	F(4,155)=0.1 (97%)	F(4,155)=1.6 (18%)
ARCH	F(4,155)=1.2 (29%))	F(4,155)=0.5 (71%)
Markov	F(4,155)=0.9 (47%)	F(4,155)=1.7 (15%)

Standard errors based on conditional scores are given in parenthesis. The displayed misspecification tests are also based on conditional scores and are F versions from papers by Newey (1985), Tauchen (1985), and White (1987). They are suggested in Hamilton (1996) for the Markov model. The significance level is given in parenthesis. None of the displayed diagnostics indicate any specification problems, although it is an open issue how these tests perform in finite samples for HMM.

There is a close similarity between the state dependent covariances for the disturbances in this paper and that in Blix (1997 a, b), which examined the short end of the yield curve (three and six month treasury bills). In both cases the high variance state (state one) has negatively correlated disturbances, while in the other state the disturbances are virtually uncorrelated. This appears to be a robust feature of the specification across the maturity spectrum.

It also appears to be a robust feature across more restrictive versions of model 1. As can be seen from the point estimates and the standard errors, neither the drift nor the autoregressive terms differ widely - with the exception of the mean-reversion for the spread. This suggests estimating HMM with several parameters constant across states. Table 2 displays the log-likelihood values of models with less parameters depending on the state and corresponding LR tests for the most parsimonious specification. The only hypothesis that can be rejected is one in which the covariance term is constant across states. In particular, we cannot reject the hypothesis that both the drift and the AR terms are state invariant. The estimates of these more restrictive HMM are not displayed, but they all have the above discussed feature that the high variance state has negatively correlated disturbances, while being uncorrelated in the other state.

Table 2: more parsimonious HMM for the US

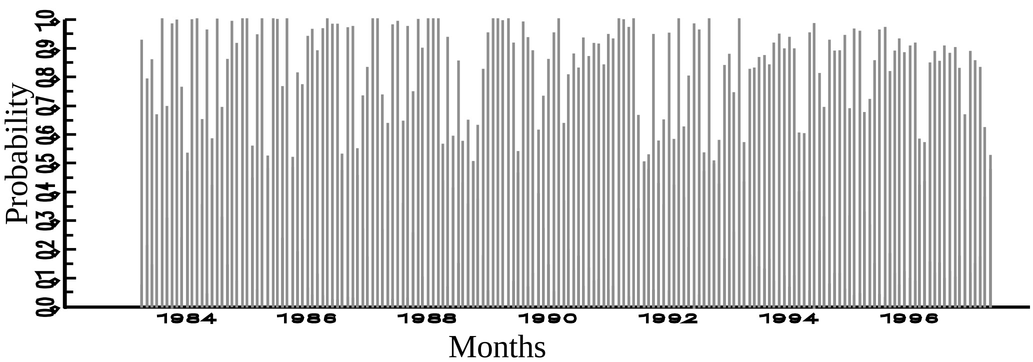
Model	State Dependence	Log-likelihood
1	$\mu_{s_t}, B_{s_t}, \Omega_{s_t}$	$L_1(\hat{\theta}) = 5.4$
2	μ_{s_t}, Ω_{s_t}	$L_2(\hat{\theta}) = -0.16$
3	Ω_{s_t}	$L_3(\hat{\theta}) = -1.4$
4	μ_{s_t}	$L_4(\hat{\theta}) = -30.0$

Model Selection

$$\text{LR}(8) = 2(L_1 - L_2) = 2(5.4 + 0.16) = 11.2 \ (19\%)$$
$$\text{LR}(10) = 2(L_1 - L_3) = 2(5.4 + 1.4) = 13.6 \ (19\%)$$
$$\text{LR}(11) = 2(L_1 - L_4) = 2(5.4 + 30.0) = 70.8 \ (0\%)$$

In table 2, $L_\tau(\hat{\theta})$ denotes the MLE of the parameters for model τ. The displayed log-likelihood ratios are distributed χ^2 under the null hypothesis with degrees of freedom and the significance level given in the parentheses. The tests are clearly not independent, but are still suggestive that the most important part of the specification is the state dependent covariance. The results thus indicate that state dependence in the drift and AR part is not statistically significant. Nevertheless, we are going to use the most general specification in model 1: all these HMM give essentially the same result both for the expectations hypothesis, the conditional premium, as well as the state probabilities, plotted in figure 3a.

Figure 3a. $\Pr(s_t = 1|y_t)$ US, model 1



4.1.2 Estimates for Sweden

The corresponding estimates for Sweden are displayed in table 3. The broad picture presented here is similar to the US: the disturbances are negatively correlated in the high-volatility state, while uncorrelated in the other; the standard errors are fairly large; and it appears that the most important part in the specification is the state dependent covariance matrix (see table 4 below).

Table 3. Model 1, Sweden 1984\03-1997\03.

$$L_1(\hat{\theta}) = -189.7, \quad \delta = 0.992, \quad \mathrm{E}[S_t] = 0.35, \quad \mathrm{E}[\Delta r_t] = -0.04, \quad P = \begin{pmatrix} \underset{(0.11)}{0.53} & \underset{(0.07)}{0.28} \\ \underset{(0.11)}{0.47} & \underset{(0.07)}{0.72} \end{pmatrix},$$

State 1

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.22)}{0.08} \\ \underset{(0.19)}{0.06} \end{pmatrix} + \begin{pmatrix} \underset{(0.37)}{0.48} & \underset{(0.46)}{0.87} \\ \underset{(0.33)}{0.06} & \underset{(0.39)}{0.92} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.32)}{0.01} & \underset{(0.54)}{-0.57} \\ \underset{(0.29)}{0.09} & \underset{(0.45)}{-0.11} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix},$$

$$\Omega_1 = \begin{pmatrix} \underset{(0.289)}{1.266} & \underset{(0.242)}{-1.087} \\ \underset{(0.242)}{-1.087} & \underset{(0.218)}{1.061} \end{pmatrix}$$

State 2

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.05)}{-0.11} \\ \underset{(0.05)}{-0.03} \end{pmatrix} + \begin{pmatrix} \underset{(0.10)}{0.36} & \underset{(0.11)}{0.37} \\ \underset{(0.10)}{-0.14} & \underset{(0.11)}{0.77} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.05)}{0.06} & \underset{(0.11)}{-0.31} \\ \underset{(0.06)}{-0.21} & \underset{(0.12)}{0.17} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix}$$

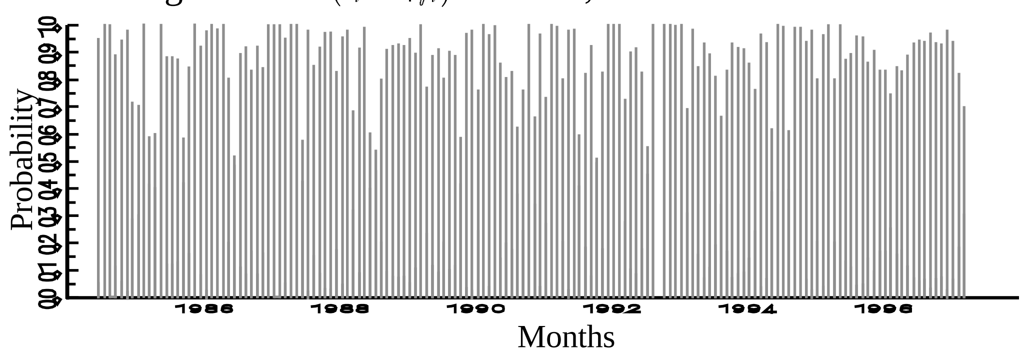
$$\Omega_2 = \begin{pmatrix} \underset{(0.022)}{0.102} & \underset{(0.022)}{-0.080} \\ \underset{(0.022)}{-0.080} & \underset{(0.028)}{0.123} \end{pmatrix}.$$

Misspecification Tests

	Equation 1	Equation 2
Autocorrelation	F(4,138)=1.7 (16%)	F(4,138)=0.7 (61%)
ARCH	F(4,138)=2.7 (3%)	F(4,138)=2.1 (8%)
Markov	F(4,138)=1.2 (31%)	F(4,138)=0.8 (50%)

The estimates differ in some respects though. The spread has much weaker mean-reversion than for the US; the Markov probability (p_{11}) of returning to state one is much lower; and the ARCH test is borderline significant in equation 1.

Figure 3b. $\Pr(s_t = 1 | \mathcal{Y}_t)$ Sweden, model 1



Analogously to the table presented for the US, table 4 shows the MLE, which are denoted by $L_\tau(\hat{\theta})$ for model τ, and LR tests for less state-dependence. The tests are distributed χ^2 under the null hypothesis with degrees of freedom and the significance level in parenthesis. Although the tests are not independent, similarly to the US they suggest that the most important part of the specification is Ω_{s_t}.

Table 4: more parsimonious HMM for Sweden

Model	State Dependence	Log-likelihood
1	$\mu_{s_t}, B_{s_t}, \Omega_{s_t}$	$L_1(\hat{\theta}) = -189.7$
2	μ_{s_t}, Ω_{s_t}	$L_2(\hat{\theta}) = -196.9$
3	Ω_{s_t}	$L_3(\hat{\theta}) = -197.7$
4	μ_{s_t}	$L_4(\hat{\theta}) = -227.8$

Model Selection

$$\text{LR}(8) = 2(L_1 - L_2) = 2(196.9 - 189.7) = 14.4\ (7.2\%)$$
$$\text{LR}(10) = 2(L_1 - L_3) = 2(197.7 - 189.7) = 16.0\ (10\%)$$
$$\text{LR}(11) = 2(L_1 - L_4) = 2(227.8 - 189.7) = 76.2\ (0\%)$$

4.2 Evaluating the Restricted Model

In this section we consider various ways in which to test and evaluate the model under the EH restrictions. We will use both formal hypothesis testing and informal goodness of fit measures.

The first question concerns the form of the EH restrictions. From (16) these restrictions in model 1 are given by

$$R\,\text{vec}\,b = r, \tag{22}$$

where $R = I_{npq} \otimes \delta(1 \quad 1)$, and $r = \text{vec}\,e_2\text{'}JC_{np}$. The restrictions in (22) can be written as

$$B_\tau^{(k)} = \begin{cases} \begin{pmatrix} B_\tau^{(1,1,1)} & B_\tau^{(1,2,1)} \\ -B_\tau^{(1,1,1)} & 1/\delta - B_\tau^{(1,2,1)} \end{pmatrix} & \text{for } k = 1 \\ \begin{pmatrix} B_\tau^{(1,1,k)} & B_\tau^{(1,2,k)} \\ -B_\tau^{(1,1,k)} & -B_\tau^{(1,2,k)} \end{pmatrix} & \text{for } k = 2,3,\cdots,p \end{cases} \tag{23}$$

where $B_\tau^{(k)}$ is the $k{:}th$ lag matrix in state τ, and $B_\tau^{(i,j,k)}$ is the $i,j{:}th$ element of $B_\tau^{(k)}$. Note that the restrictions follow the pattern given for the single regime VAR in Warne (1990, p 71) and are identical to Campbell & Shiller (1987). If we want to test the EH restrictions for model 2 in which $B_{s_t} = B$, we simply remove the τ subscript in (23). As the restrictions are linear, re-estimating the VAR:s under those restrictions is not particularly computationally demanding.

Table 5

Model	State Dependence	$\delta = 0.993$ US	$\delta = 0.991$ Sweden
1	$\mu_{s_t}, B_{s_t}, \Omega_{s_t}$	$L_1(\tilde{\theta}) = -15.5$	$L_1(\tilde{\theta}) = -208.2$
2	μ_{s_t}, Ω_{s_t}	$L_2(\tilde{\theta}) = -18.3$	$L_2(\tilde{\theta}) = -210.3$
3	Ω_{s_t}	$L_3(\tilde{\theta}) = -19.5$	$L_3(\tilde{\theta}) = -213.1$

LR tests of Expectations Hypothesis with k degrees of freedom

Model	k	US	Sweden
1	8	41.8 (0%)	36.6 (0%)
2	4	36.2 (0%)	26.8 (0%)
3	4	36.2 (0%)	40.8 (0%)

Table 5 displays log-likelihood values of the ML estimates under the EH restrictions, which are indicated by the use of tildes, and the corresponding LR tests for the models in tables 2 and 4. The estimates for model 1 (only) are given in the appendix. The discount factors are computed from $\delta = \left(1 + \bar{R}\right)^{-1}$; they are almost the same for both countries, and correspond to a monthly discount rate of about 0.7 %, or 8.7 % on a yearly basis. All the tests are highly significant, and strongly reject the EH.

Second, we investigate whether this strong rejection of the EH depends on our choice of discount factor. Instead of changing a single discount factor, suppose the discount factor depends on the current state (only). Specifically, the standard discount factor of y_{t+j} at time t is δ^j, but instead we use $\prod_{\tau=1}^{j}\delta_{s_{t+\tau}}$, where the δ_τ $\tau=1,\cdots,q$ are exogenous. This gives more flexibility to the EH test, which is potentially important.

To find the EH restrictions in this setting we need a slight modification of (22). It was shown in Blix (1997 a,b) that such a generalisation is straightforward where we replace R in (22) by

$$R_\delta = \delta^{(np)} \otimes \begin{pmatrix}1 & 1\end{pmatrix} \tag{24}$$

where $\delta^{(\tau)} = \delta \otimes I_\tau$ and $\delta = \operatorname{diag}\left(\delta_1,\cdots,\delta_q\right)$. Corresponding to (23), the restrictions can also be written as

$$B_\tau^{(k)} = \begin{cases} \begin{pmatrix} B_\tau^{(1,1,1)} & B_\tau^{(1,2,1)} \\ -B_\tau^{(1,1,1)} & 1/\delta_\tau - B_\tau^{(1,2,1)} \end{pmatrix} & \text{for } k=1 \\ \begin{pmatrix} B_\tau^{(1,1,k)} & B_\tau^{(1,2,k)} \\ -B_\tau^{(1,1,k)} & -B_\tau^{(1,2,k)} \end{pmatrix} & \text{for } k=2,3,\cdots,p. \end{cases} \tag{25}$$

Since we want to try a large number of different values for δ_τ, it is inconvenient to re-estimate the model for every different set of discount factors. Instead, we use a Wald test given by

$$W = T\left(R_\delta\beta - r\right)'\left[RQR'\right]^{-1}\left(R_\delta\beta - r\right) \sim \chi^2(npq) \text{ on } \mathrm{H}_0\text{: } R_\delta\beta - r = 0, \tag{26}$$

where Q is the covariance matrix of the disturbances. This test has the advantage of relying on the unrestricted estimates, but the potential disadvantage that its finite sample properties depend on the way the restrictions are written, as discussed in Gregory and Veall (1985).

Is the rejection of the EH robust to this extension? As a benchmark case for model 1 we use the single regime value of δ, which gives a Wald statistic of 38.5 distributed with 8 degrees of freedom; this is highly significant at conventional levels: the 1% level is W(8)=20.1. No matter how much we change $\delta_\tau \in (0,1)$, we do not change the value of the Wald test in (26) by more than a very small amount: the minimum value we found was 37.5 where $\delta_1 = 0.999$ and $\delta_2 = 0.969$. For Sweden the statistic is similarly robust to changes in the discount factor.

A third question of interest is whether despite the strong statistical rejection of the EH restrictions, the EH still has some merit in terms of goodness of fit. Following Campbell & Shiller (1987) we compute the ex ante optimal unrestricted forecast given by

$$S_t^* \equiv \mathrm{E}\left[\sum_{j=1}^{\infty} \Delta r_{t+j} \middle| \mathcal{Y}_t\right] = b^* \tilde{Y}_t, \tag{27}$$

where b^* is defined in (18). The extent to which b^* is close to $e_2{}'JC_{np}$ measures how close the unrestricted estimates are to fulfilling the EH; the correlation between S_t and S_t^* gives a measure of goodness of fit to complement formal test statistics.

Figures 4a and 4b display the forecast spread compared to the actual spread of the US and Sweden respectively for model 1. In the forecast for Sweden we have deleted the extreme observation in September 1992; in the plot we have inserted missing values for that observation, both for the forecast and the actual spread.

Figure 4a. Ex ante Optimal Forecast, US model 1

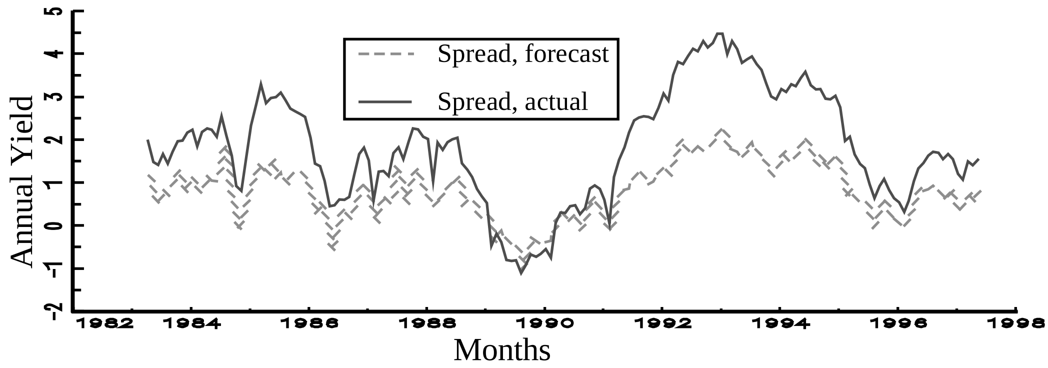


Figure 4b. Ex ante Optimal Forecast, Sweden model 1

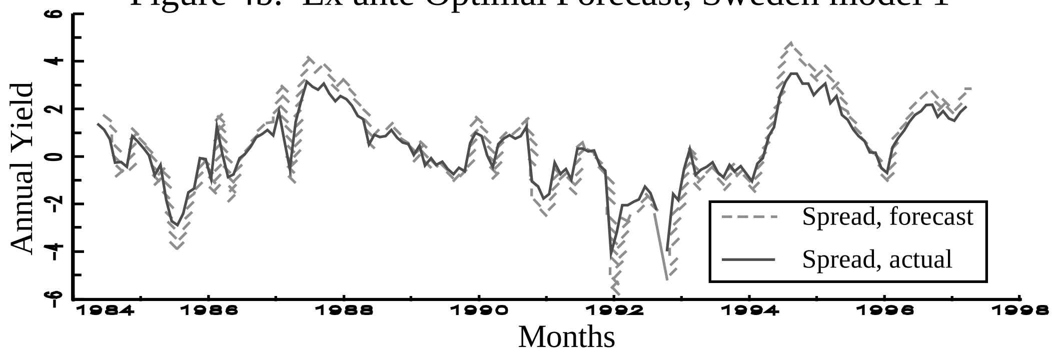


Table 6 shows more formally the visual results in figures 4a and 4b. We find that the series are highly correlated, although the variance ratios are not close to unity. Although the correlations for Sweden are at par with the US, the forecast of the spread for Sweden tracks the actual spread much more closely.

Table 6

Model	$\mathrm{Corr}(S_t, S_t^*)$		$\mathrm{Var}[S_t] / \mathrm{Var}[S_t^*]$	
	US	Sweden	US	Sweden
1	0.956	0.995	3.80	0.56
2	0.980	0.977	3.21	1.60
3	0.980	0.970	3.25	2.14

For the US, model 2 gives the best results, but the other models are very similar. For Sweden, model 1 gives the best fit, but also has a higher variance of the forecast than the actual spread. Overall, predictive performance is good for the EH.

Even though the EH is statistically rejected, we can thus view the EH as a useful approximation. In particular, the magnitude of the conditional premium, displayed in table 7, is still of interest. This table gives the premium ψ_τ in (17) under the null hypothesis of the EH, using the expressions for λ in (18) and (21) for models 1 and 2 respectively. Note that since the premium is defined under the null only, all the parameters used have been estimated subject to the restrictions in (23). Here, by contrast, we find that the results are highly sensitive to the choice of δ. Using the benchmark values for δ above, we obtain the (implausibly) large premia displayed in table 7; for a monthly rate of 1.1 % instead, which corresponds to a 14 % yearly rate, we obtain smaller values.

Table 7: Conditional Premium

		Benchmark 0.7 % monthly		1.1 % monthly		2.0 % montly	
δ:		0.993	0.991	0.989	0.989	0.980	0.980
Model	State	US	Sweden	US	Sweden	US	Sweden
1	1	4.99	6.02	3.78	3.86	2.89	2.18
	2	4.98	6.30	3.77	4.15	2.88	2.48
2	1	4.99	5.98	3.78	3.86	2.89	2.19
	2	4.98	6.26	3.77	4.14	2.88	2.47

The extreme sensitivity of the premium with respect to the discount factor and the large size of the premium are perhaps not surprising. One way to see where these features come from is to take unconditional expectations of (8), and we obtain

$$\mu_S = \frac{\delta}{1-\delta}\mu_{\Delta r} + \psi, \qquad (28)$$

where $\mu_S = \mathrm{E}[S_t]$, $\mu_{\Delta r} = \mathrm{E}[\Delta r_t]$, and $\psi = \mathrm{E}[\psi_{s_t}]$. As $\delta \to 1$, the factor multiplying $\mu_{\Delta r}$ becomes very large, and since $\mu_{\Delta r}$ is negative (see tables 1 and 3) the premium becomes large. For the US, the factor multiplying $\mu_{\Delta r}$ is about 145; for Sweden, 119.

The sensitivity of the premium with respect to δ and the large standard errors on many of the estimates require that the results should be interpreted with caution. In particular, inferences about the level of the premium appear unsafe. Nevertheless, we may draw two conclusions. First, the US premium is equal across regimes for all models considered and this appears to be a robust feature. Second, for Sweden there is a difference of about 30 basis points in the premium between the two states for all models considered. Thus, only for Sweden does the regime dependence in the drift and AR part contribute to different premia across states. Overall, it appears that this form of state dependence in the premium neither leads to non-rejection of the EH nor improves the fit of the EH over and above that achieved by having state dependence in the covariance of the disturbances.

5 Concluding Remarks

In this paper we have tested the PVM when allowing for stochastic switches in regimes. We have modeled the term premium as depending on the current state. This approach gives a parsimonious representation of non-linearities in the data, and allows the term premium to differ across states under the null hypothesis of the EH.

This more flexible specification does not, however, resurrect the EH. It is strongly rejected for both Sweden and the US with conventional LR tests. An extended version of the PVM with state dependent discount rate did not change this result: no matter how much the discount rate was changed inside the unit interval, the Wald statistic only changed very slightly. But despite these statistical rejections we found that the EH provides a good approximation to the data, similar to the finding in Campbell & Shiller (1987).

Finally, we found that the most important feature of the estimated VAR:s for Sweden and the US to be the state dependent covariance of the disturbances; all other state dependence was statistically rejected. This resulted in term premia that did not differ much across states. This does not, however, provide evidence for the constant premium assumption in the literature. The estimated premium is derived under the hypothesis that the EH *holds* and is undefined under the alternative.

Appendix A

Derivation of λ

Recall that

$$\psi_\tau = -e_1'\left[a\sum_{j=1}^{\infty}(\delta P)^j + b\sum_{j=2}^{\infty}\delta^j\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right]e_\tau. \tag{A.1}$$

The first term on the RHS gives

$$\begin{aligned} e_1'a\delta P\overline{P}_1^{-1}e_\tau &= \delta\,\mathrm{tr}\left[e_1'aP\overline{P}_1^{-1}e_\tau\right] \\ &= \delta\, e_\tau'\left(I_q\otimes e_1'a\right)\mathrm{vec}\left[P\overline{P}_1^{-1}\right] \\ &= \delta\left[E_1'\left(I_q\otimes a\right)\mathrm{vec}\left[P\overline{P}_1^{-1}\right]\right]'e_\tau, \end{aligned} \tag{A.2}$$

since a scalar is equal to its trace, and $\mathrm{tr}\, ABCD = (\mathrm{vec}\, D')'(C'\otimes A)\mathrm{vec}\, B$, see for example Magnus and Neudecker (1988). Similarly, the second term is

$$e_\tau'E_1'\left(I_q\otimes b\right)\sum_{j=2}^{\infty}\delta^j\,\mathrm{vec}\, S_{j-2}^*, \tag{A.3}$$

where $S_\tau^* \equiv \sum_{m=0}^{\tau}\Phi^m\Psi P^{\tau+1-m}$. Note that $S_{j-2}^*P - \Phi S_{j-2}^* = \Psi P^j - \Phi^{j-1}\Psi P$, and so $\left[P_{npq}' - \Phi_q\right]\mathrm{vec}\, S_{j-2}^* = P_{npq}'\left[\left(P_{npq}'\right)^{j-1} - \Phi_q^{j-1}\right]\mathrm{vec}\Psi$. Thus,

$$\mathrm{vec}\, S_{j-2}^* = D^{-1}P_{npq}'\left[\left(P_{npq}'\right)^{j-1} - \Phi_q^{j-1}\right]\mathrm{vec}\Psi. \tag{A.4}$$

Substitute (A.4) into (A.3), which yields

$$e_\tau'E_1'\left(I_q\otimes b\right)D^{-1}\delta P_{npq}'\sum_{j=2}^{\infty}\left[\left(\delta P_{npq}'\right)^{j-1} - \left(\delta\Phi_q\right)^{j-1}\right]\mathrm{vec}\,\Psi. \tag{A.5}$$

Summing and transposing we obtain

$$\delta^2\left[E_1'\left(I_q\otimes b\right)D^{-1}P_{npq}'\left(P_{npq}'\tilde{P}_1^{-1} - \Phi_q\tilde{\Phi}_1^{-1}\right)\mathrm{vec}\,\Psi\right]'e_\tau. \tag{A.6}$$

Using (A.2) and (A.6) together gives the desired result.

Appendix B

Data Definitions and Sources:

Data for US

r_t = 1 m Euro Market USD rate, Sveriges Riksbank, 1963/07-1997/3.
R_t = 30 y constant treasury maturity rate, FRED database, 1977/2-1997/3.

Data for Sweden

r_t = 30 day rate, Sveriges Riksbank, 1983/01-1997/03.
R_t = 5 year treasury bill, Sveriges Riksbank, 1984/03-1997/03.

Appendix C

Estimates of model 1 under EH restrictions

Table C1. Model 1, US 1982 / 11 - 1997 / 3.

$$L_1(\tilde{\theta}) = -15.5, \quad \delta = 0.993, \quad P = \begin{pmatrix} \underset{(0.08)}{0.79} & \underset{(0.07)}{0.17} \\ \underset{(0.08)}{0.21} & \underset{(0.07)}{0.83} \end{pmatrix},$$

State 1

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.09)}{-0.15} \\ \underset{(0.08)}{0.11} \end{pmatrix} + \begin{pmatrix} \underset{(0.19)}{0.06} & \underset{(0.24)}{0.12} \\ \underset{(0.19)}{-0.06} & \underset{(0.24)}{0.89} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.19)}{0.20} & \underset{(0.24)}{-0.07} \\ \underset{(0.19)}{-0.20} & \underset{(0.24)}{0.07} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix},$$

$$\Omega_1 = \begin{pmatrix} \underset{(0.004)}{0.207} & \underset{(0.003)}{-0.150} \\ \underset{(0.003)}{-0.150} & \underset{(0.003)}{0.150} \end{pmatrix}$$

State 2

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.03)}{-0.04} \\ \underset{(0.04)}{0.01} \end{pmatrix} + \begin{pmatrix} \underset{(0.08)}{0.43} & \underset{(0.06)}{0.26} \\ \underset{(0.08)}{-0.43} & \underset{(0.06)}{0.75} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.06)}{0.04} & \underset{(0.06)}{-0.22} \\ \underset{(0.06)}{-0.04} & \underset{(0.06)}{0.22} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix}$$

$$\Omega_2 = \begin{pmatrix} \underset{(0.02)}{0.01} & \underset{(0.03)}{0} \\ \underset{(0.03)}{0} & \underset{(0.01)}{0.07} \end{pmatrix}.$$

Table C2. Model 1, Sweden, 1984 / 03 - 1997 / 03, excl. Sep. 1992

$$L_1(\tilde{\theta}) = -208.2, \quad \delta = 0.991, \quad P = \begin{pmatrix} \underset{(0.08)}{0.76} & \underset{(0.06)}{0.18} \\ \underset{(0.08)}{0.24} & \underset{(0.06)}{0.82} \end{pmatrix},$$

State 1

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.17)}{0.05} \\ \underset{(0.15)}{0.02} \end{pmatrix} + \begin{pmatrix} \underset{(0.32)}{0.16} & \underset{(0.39)}{0.41} \\ \underset{(0.32)}{-0.16} & \underset{(0.39)}{0.59} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.18)}{0.02} & \underset{(0.39}{-0.17} \\ \underset{(0.18)}{-0.02} & \underset{(0.39)}{0.17} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix},$$

$$\Omega_1 = \begin{pmatrix} \underset{(0.289)}{1.266} & \underset{(0.242)}{-1.087} \\ \underset{(0.242)}{-1.087} & \underset{(0.218)}{1.061} \end{pmatrix}$$

State 2

$$\begin{pmatrix} \Delta r_t \\ S_t \end{pmatrix} = \begin{pmatrix} \underset{(0.04)}{-0.14} \\ \underset{(0.04)}{-0.02} \end{pmatrix} + \begin{pmatrix} \underset{(0.13)}{-0.03} & \underset{(0.13)}{0.01} \\ \underset{(0.13)}{0.03} & \underset{(0.13)}{1.00} \end{pmatrix} \begin{pmatrix} \Delta r_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \underset{(0.06)}{0.10} & \underset{(0.13)}{0.05} \\ \underset{(0.06)}{-0.10} & \underset{(0.13)}{-0.05} \end{pmatrix} \begin{pmatrix} \Delta r_{t-2} \\ S_{t-2} \end{pmatrix}$$

$$\Omega_2 = \begin{pmatrix} \underset{(0.02)}{0.09} & \underset{(0.02)}{-0.06} \\ \underset{(0.02)}{-0.06} & \underset{(0.02)}{0.10} \end{pmatrix}.$$

References

Blix, Mårten (1997a), "Rational Expectations in a VAR with Markov Switching", chapter 2 in *Rational Expectations and Regime Shifts in Macroeconometrics*, PhD dissertation at the Institute for International Economic Studies, Stockholm University. Akademitryck: Edsbruk.

Blix, Mårten (1997b), "Rational Expectations in a VAR with Markov Switching", *Sveriges riksbank working paper*, No 42.

Campbell, John Y. and Robert J. Shiller (1987), "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, 95, pp. 1062-1088.

Driffill, John (1992), "Changes in Regime and the Term Structure: A note", *Journal of Economic Dynamics and Control*, 16, pp. 165-173.

Garcia, René (1997), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *mimeo*, Université de Montréal.

Gregory, Allan and Michael Veall (1985), "Formulating Wald Tests of Nonlinear Restrictions", *Econometrica*, Vol. 53, pp. 1465-68.

Hamilton, James 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, pp. 385-423.

Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle", *Econometrica*, Vol. 57, No 2, pp. 357-384.

Hamilton, James D. (1990), "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, 45, pp. 39-70.

Hamilton, James D. (1994), *Time Series Analysis*, Chapter 22, Princeton University Press: Princeton, New Jersey.

Hamilton, James D. (1996), "Specification Testing in Markov-Switching Time Series Models", *Journal of Econometrics*, 70, Pp. 127-157.

Holst, Ulla, Georg Lindgren, Jan Holst, and Mikael Thuvesholmen (1994), "Recursive Estimation in Switching Autoregressions with a Markov Regime", *Journal of Time Series Analysis*, 15, No 5, pp. 489-506.

Kugler, Peter (1996), "The term structure of interest rates and regime shifts: Some empirical results", *Economic Letters*, 50, pp. 121-126.

Lindgren, Georg (1978), "Markov Regime Models for Mixed Distributions and Switching Regressions", *Scandinavian Journal of Statistics*, 5, pp. 81-91.

Magnus, Jan R. and Heinz Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics* (John Wiley: Chichester).

Newey, Whitney K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, Vol. 53, No 5, pp. 1047-1070.

Rydén, Tobias (1995), "Estimating the Order of Hidden Markov Models", *Statistics*, 26, pp. 345-354.

Shiller, Robert J. (1979), "The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure", *Journal of Political Economy*, 87, pp. 1190-1219.

Shiller, Robert J., John Y. Campbell, and Kermit L. Schoenholtz (1983), "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates", *Brookings Papers on Economic Activity*, No 1, pp. 173-217.

Shiller, Robert J. (1990), "The Term Structure of Interest Rates" in Handbook of Monetary Economics, Vol. 1 (North Holland: Amsterdam), pp. 627-722.

Sola, Martin and John Driffill (1994), "Testing the Term Structure of Interest Rates using a Stationary Vector Autoregression with Regime Switching", *Journal of Economic Dynamics and Control*, 18, pp. 601-628.

Tauchen, George (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics,* 30, pp. 415-443.

Warne, Anders (1990), "Stationarity, Cointegration and a Rational Expectations Model of the Term Structure" in *Vector Autoregressions and Common Trends in Macro and Financial Economics,* PhD Dissertation Stockholm School of Economics, pp. 62-102.

Warne, Anders (1996), "Autocovariance Functions and Maximum Likelihood in a VAR Model under Markov Switching", *Mimeo*, Institute for International Economic Studies, Stockholm University, Sweden.

White, Halbert (1987), "Specification Testing in Dynamic Models" in Truman F. Bewley, ed., *Advances in Econometrics*, Fifth World Congress, Vol. 1, Cambridge University Press: Cambridge.