# Rational Expectations in a VAR with Markov Switching

Mårten Blix*

First version: October 1996
This version: 1997/05/09

Comments Welcome

**Abstract**

This paper shows how a well known class of rational expectations hypotheses using linear vector autoregressions (VAR:s) can be extended to allow for unobservable Markov switching. The regime shift model used falls into the general framework of Hamilton (1990), but differs to the *centered* model actually implemented by Hamilton and others. The model here has the advantage that it is easier to estimate, and the intuitive appeal that the state dependence is symmetric. The contribution of the paper is to derive testable restrictions implied by rational expectations, which are linear when the forecast horizon is infinite. The restrictions on the autoregressive parameters are the same as those that appear in the *centered* model. As an illustration, we duplicate a test of the expectations hypothesis (EH) in Sola & Driffill (1994) on 3 and 6 month US bills on quarterly data, and find that their results may be fragile.

JEL Classification No: C12, C32.

Keywords: VAR, Markov chain, regime switching, rational expectations, expectations hypothesis.

---

* Institute for International Economic Studies, S-106 91 Stockholm, Sweden. Tel: +46 8 163058, fax: +46 8 161443. Email: Blixm@iies.su.se. I am much indebted to Anders Warne for helpful comments and for generously supplying his Gauss code for the EM algorithm. I would also like to thank Tobias Rydén and Lars E.O. Svensson for valuable suggestions. Financial support from the Jan Wallander and Tom Hedelius foundation is gratefully acknowledged.

## 1 Introduction

Many economic models postulate relationships between currently observable variables and expectations of future variables. Given a parametric (or semi-parametric) form for the evolution of these variables, rational expectations (RE) typically imply restrictions on the parameters of the statistical model. For example, the present value model considered by Shiller (1979) posits that the yield of a consol should equal discounted expectations of changes in future short rates; when the dynamics of these variables are driven by a vector autoregression (VAR), the hypothesis of rational expectations implies restrictions on the parameters of the VAR (Campbell and Shiller 1987).

This paper examines a class of rational expectations hypotheses that is well known in a VAR setting, and extends it to a VAR with unobservable Markov switching to $q$ different states. The regime shift model used falls into the general framework of Hamilton (1990,1994), but differs to the *centered* model implemented by Hamilton and others. The model here has the advantage that it is generally easier to estimate, and that the state dependence is symmetric (in a sense that will be made clear below).

The contribution of the paper is to derive testable restrictions implied by rational expectations when there are switching regimes; the restrictions are non-linear, but are presented in compact matrix form allowing easy implementation. In the important case of infinite horizon models, however, they are linear - and do not involve the Markov transition probabilities. Moreover, the restrictions on the autoregressive parameters are the same as those that would appear from the *centered* regime model, although for the drift term they differ. Since most interest lies with testing restrictions on the autoregressive parameters, the results in this paper may have a wider appeal.

Some of the renewed interest in models with regime shifts since Hamilton (1988) may stem from the failure - or statistical rejection - of simple linear models. Markov models can provide an appealing alternative. First, a model with Markov switching may provide a better characterisation of data. In other words, it may provide a parsimonious way to express complicated dynamics, which might otherwise require a ARIMA model with long lags, an issue briefly discussed in Hamilton (1989). With this view the states do not necessarily have any explicit "interpretation", such as "high" or "low" risk, but this might also be justified, as in Warne (1996).

Second, discrete states may be a useful tool in economic modelling, for which Markov switching provides a first step towards empirical work. Even when discrete states cannot easily be mapped onto real environments, they may provide an (arbitrary) approximation to some continuous phenomenon.

Finally, the question arises whether or not RE hypotheses hitherto rejected in single-regime models will be resurrected in models with randomly switching coefficients. For example, Hamilton (1988) considers a regime shift model for the long end of the term structure. He finds that the single regime model does not fit the data and that RE can no longer be rejected in his regime shift model. Similar conclusions are reached by Sola & Driffill (1994) for the short end of the maturity spectrum. As an illustration of the methods developed in this paper, we re-examine the Sola & Driffill (1994) non-rejection of the EH and find that it may be fragile.

The rest of this paper is outlined as follows. The next section formally introduces the VAR with Markov switching. Section 3 formalises the class of hypotheses that are considered, and provides some examples. Although the examples can be skipped, they are used throughout the paper to illustrate the main results of the paper, given in section 4, where the restrictions on the parameters are derived and discussed. Section 5 considers the same restrictions but in a model with state dependent discount rate. Section 6 discusses statistical tests of the restrictions, and also gives the required derivatives. Section 7 illustrates the use of the methods with a test the expectations hypothesis of the term structure, using 3 and 6 month US bills. Section 8 makes some concluding remarks.

## 2 A VAR with Markov Switching

The model we consider 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, $\mu_{s_t}$ is the vector of state dependent intercepts, and $\Omega_{s_t}$ 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. With the usual notation, we define the Markov transition probabilities as $p_{ij} = \text{pr}[s_t = j | s_{t-1} = i]$, and collect them into the matrix

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

where $p_{\tau q} = 1 - \sum_{j=1}^{q-1} p_{\tau j}$, 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 ergodic (stationary) probabilties are defined by the property that $P\pi = \pi$. If each column of the transition matrix is equal to $\pi$, we have a serially uncorrelated MC, in which the probability of staying in a particular state is the same as the probability of returning to it from all the other states. Such a transition matrix has rank one.

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

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

This model is conveniently cast in companion form, in which a VAR($p$) is compactly re-written as the VAR(1) system

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

where

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

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

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

A sufficient condition for weak stationarity from Karlsen (1990) is that the largest eigenvalue of the $(np)^2 q \times (np)^2 q$ matrix

$$B^* = \begin{bmatrix} (B_1 \otimes B_1) p_{11} & \cdots & (B_1 \otimes B_1) p_{q1} \\ \vdots & & \vdots \\ (B_q \otimes B_q) p_{1q} & \cdots & (B_q \otimes B_q) p_{qq} \end{bmatrix} \quad (7)$$

should be less than unity. The results in Warne (1996) indicate that it may also be a necessary condition.

Finally, let us comment on the differences between the regime shift model in (1) , which we will label the Warne model, and those used in Hamilton (1988,1989), which might be referred to as *centered* regime shift models. As discussed in Warne (1996), although both belong to the general class of regime switching models in Hamilton (1990,1994), they are not nested. To see this, consider the fairly general *centered* regime shift model given by

$$y_t - \mu_{s_t} = B_{s_t}^{(1)}(y_{t-1} - \mu_{s_{t-1}}) + \cdots + B_{s_t}^{(p)}(y_{t-p} - \mu_{s_{t-p}}) + \varepsilon_t . \quad (8)$$

For a given lag length $p$ and number of regimes $q$, (1) and (8) differ in the drift term. Moreover, the *centered* model is non-linear in some parameters even after conditioning on current and past states. By contrast, the Warne model is linear after conditioning on the current state, and is therefore much easier to estimate. Moreover, the regime dependence in the *centered* model is asymmetric in that some parameters depend on the current state, while others depend both on the current and past states. Even if the $i{:}th$ lag of (8) is replaced by $B_{s_{t-i}}^{(i)}$ - or by a constant as in Hamilton (1988, 1989) - the regime dependence is still asymmetric in the sense that some parameters depend only on the current state, while others depend only on past states.

How might we choose between them? It is shown in Warne (1996) that the (un)conditional autocovariances of (8) are the same as those for (1) under certain conditions. Both models allow for rich dynamics, but *a priori* it is not clear which is more suitable to a given economic model, a question left to future research. Note, however, that the model with $B_{s_t}^{(i)} = B^{(i)}$ has much more restrictive dynamics than the Warne model. Moreover, from a practical viewpoint, the Warne model is easier to handle. But for the purposes of this paper, the choice of using (1) or (8) does not matter: the restrictions on the autoregressive parameters are the same; conditional forecasts from the two models differ *only* in the drift term.

## 3 Rational Expectations Hypotheses

The question that will be pursued here is how to formulate rational expectations restrictions of the form

$$\sum_{j=0}^{k} N_j \, \mathrm{E}\left[ y_{t+j} \middle| \mathcal{Y}_t, s_t \right] = \lambda_{s_t}, \tag{9}$$

where $N_j$ is a $s \times n$ selection matrix, and $\lambda_{s_t}$ is a $s \times 1$ vector that depends on the current state only. Since by assumption we do not observe the current state directly, we take expectations of (9) conditional on only the observable information $\mathcal{Y}_t$ to obtain

$$\sum_{j=0}^{k} N_j \, \mathrm{E}\left[ y_{t+j} \middle| \mathcal{Y}_t \right] = \mathrm{E}\left[ \lambda_{s_t} \middle| \mathcal{Y}_t \right], \tag{10}$$

where we have used the law of iterated expectations.

For standard (single-regime) VAR:s, the survey by Baillie (1989) discusses several RE applications in some detail. The form of the hypothesis differs only slightly from Baillie in that the RHS of (10) is a state dependent vector, but this will not make a difference for many applications (i.e. $\lambda_i = 0, \forall i \in \{1, \cdots, q\}$). To motivate the discussion below, we will briefly provide some examples of RE hypotheses that fall into the category of (10), either directly or after some suitable transformation. The examples will also serve to illustrate some possible interpretations for $\lambda_{s_t}$. Although the rest of this section can be skipped without loss of continuity for the theoretical exposition, some of the examples in this section will be used to illustrate how the results in the paper can be applied.

**Example 1: Term Structure**

One version of the linearized expectations model for discount bonds is

$$R_t^{(2)} = 0.5\mathrm{E}\left[ R_t^{(1)} + R_{t+1}^{(1)} \middle| \mathcal{Y}_t^* \right] + \psi_{s_t}^{(2)}, \tag{11}$$

where $R_t^{(i)}$ is the yield at time $t$ on a bond with maturity $t+i$, and $\psi_{s_t}^{(\tau)} (= \lambda_{s_t})$ is a premium on $\tau$ period bonds. Thus, this expectations hypothesis states that the yield of a two period bond should equal the expected value of holding two one-period bonds over the life of the two-period bond. The only non-standard feature of (11) is that the premium is assumed to be state dependent. This approach has been used in Blix (1996) to model a conditional term premium.

Subtract $R_t^{(1)}$ from both sides, and take expectations conditional on $\mathcal{Y}_t$,

$$S_t = 0.5\mathrm{E}\left[\Delta R_{t+1}^{(1)} \middle| \mathcal{Y}_t\right] + \mathrm{E}\left[\psi_{s_t}^{(2)} \middle| \mathcal{Y}_t\right], \tag{12}$$

where $S_t \equiv R_t^{(2)} - R_t^{(1)}$ is the spread between the long and the short rate. It can now be written as

$$e_1' y_t - 0.5 e_2' \mathrm{E}[y_{t+1} | \mathcal{Y}_t] = \mathrm{E}[\psi_{s_t}^{(2)} | \mathcal{Y}_t], \tag{13}$$

where $y_t = \left(S_t, \Delta R_t^{(1)}\right)'$, and $e_i$ is the $i{:}th$ column of an identity matrix of order 2.

**Example 2: Asset Pricing Models**

The present value model for stock prices states that the current price is given by the discounted value of future expected dividends, or alternatively in the form discussed in Campbell & Shiller (1987),

$$S_t = \frac{\delta}{1-\delta} \sum_{j=1}^{\infty} \delta^j \mathrm{E}\left[\Delta d_{t+j} \middle| \mathcal{Y}_t\right], \tag{14}$$

where $S_t \equiv P_t - \delta(1-\delta)^{-1} d_t$ is the spread between the stock price $P_t$ and the dividend $d_t$, and $\delta$ is the discount factor. This can be put into the framework of (10) as

$$e_1' y_t - \delta(1-\delta)^{-1} e_2' \sum_{j=1}^{\infty} \delta^j \mathrm{E}[y_{t+j} | \mathcal{Y}_t] = 0 \tag{15}$$

where $y_t = [S_t \quad \Delta d_t]'$. Here we have just ignored the premium, but it could of course be included as in example 1.

**Example 3: Uncovered Interest Rate Parity**

Engel and Hamilton (1990) consider the hypothesis of uncovered interest rate parity (UIP) in a model with Markov switching. They let

$$y_t = \begin{bmatrix} e_t^{\mathrm{G}} - e_{t-1}^{\mathrm{G}} \\ i_t^{\mathrm{US}} - i_t^{\mathrm{G}} \end{bmatrix} \tag{16}$$

where $e_t^{\mathrm{G}}$ and $i_t^{\mathrm{US}} - i_t^{\mathrm{G}}$ are the exchange rate and the interest rate differential between the US and Germany respectively. They assume that $y_t | s_t \sim \mathrm{N}\left(\mu_{s_t}, \Omega_{s_t}\right)$. The standard UIP holds that $i_t^{\mathrm{US}} - i_t^{\mathrm{G}} = \mathrm{E}\left[e_{t+1}^{\mathrm{G}} - e_t^{\mathrm{G}} \middle| \mathcal{Y}_t\right]$, which can be formulated as

$$e_2' y_t - e_1' \mathrm{E}\left[y_{t+1} \middle| \mathcal{Y}_t\right] = 0. \tag{17}$$

More complicated restrictions where $N_j$ is a matrix are also easily handled within this framework; this method of representing restrictions within a VAR framework is well known. The value added of this paper is in calculating the expectations in (10) for the regime shift VAR($p$) introduced above, which is the subject of the next section.

## 4 Rational Expectations Restrictions

The RHS of (10) is readily seen to be

$$
\begin{aligned}
\mathrm{E}\left[\lambda_{s_t}\middle|\mathcal{Y}_t\right] &= \sum_{i_0=1}^{q}\lambda_{i_0}\,\mathrm{pr}\left[s_t=i_0\middle|\mathcal{Y}_t\right] \\
&= \lambda_1\mathrm{pr}\left[s_t=1\middle|\mathcal{Y}_t\right]+\cdots+\lambda_q\mathrm{pr}\left[s_t=q\middle|\mathcal{Y}_t\right] \\
&= \lambda'\,\xi_{t|t},
\end{aligned}
\tag{18}
$$

where $\lambda=\begin{bmatrix}\lambda_1 & \cdots & \lambda_q\end{bmatrix}'$. There are at least two ways to interpret $\lambda$. First, we might want to set $\lambda_i$ arbitrarily to enunciate some a priori characteristic to that state, such as a "high" or "low" level of return, or perhaps the more standard $\lambda_i=0,\forall i\in\{1,\cdots,q\}$. Second, under certain assumptions it might be estimated as a non-linear function of the VAR parameters. As alluded to in example 1, this was the approach used in Blix (1996) to let $\lambda'\,\xi_{t|t}$ be a conditional term premium.

The next step is to consider the LHS of (10). We note that we are going to need an expression for $y_{t+j}$, and take into account expectations of future regimes. For this purpose we introduce the following lemmas.

**Lemma 4.1**

For $j\geq 2$,

$$
\begin{aligned}
y_{t+j} = \mu_{s_{t+j}} &+ J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h}B_{s_{t+j+1-m}}\right)J'\,\mu_{s_{t+j-h}} + J\left(\prod_{m=1}^{j}B_{s_{t+j+1-m}}\right)Y_t \\
&+\varepsilon_{t+j} + J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h}B_{s_{t+j+1-m}}\right)J'\,\varepsilon_{t+j-h}.
\end{aligned}
\tag{19}
$$

**Proof:** Consider $Y_{t+j}$ and substitute "backwards" until the RHS contains variables dated *t*, and then pre-multiply by $J$. □

**Lemma 4.2**

Let $e_i$ be the $i{:}th$ column of an identity matrix. For $j\geq 2$,

$$
\mathrm{pr}\left[s_{t+j}=i_j,s_{t+j-1}=i_{j-1},\cdots,s_{t+1}=i_1\middle|\mathcal{Y}_t\right]=\left(\prod_{f=0}^{j-2}e_{i_{j-f}}'\,Pe_{i_{j-f-1}}\right)e_{i_1}'\,P\xi_{t|t}. \tag{20}
$$

**Proof:** In the appendix. □

We will also use the relation

$$
\begin{aligned}
\mathrm{E}\left[y_{t+j}|\mathcal{Y}_t\right] = \sum_{i_j=1}^{q}\cdots\sum_{i_1=1}^{q}&\mathrm{E}\left[y_{t+j}\middle|s_{t+j}=i_j,s_{t+j-1}=i_{j-1},\cdots,s_{t+1}=i_1,\mathcal{Y}_t\right]\times \\
&\mathrm{Pr}\left[s_{t+j}=i_j,s_{t+j-1}=i_{j-1},\cdots,s_{t+1}=i_1\middle|\mathcal{Y}_t\right].
\end{aligned}
\tag{21}
$$

Concerning notation, let

$$B = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & B_q \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mu_q \end{bmatrix}, \tag{22}$$

where $B$ is $npq \times npq$ and $\mu$ is $nq \times q$. Also let $b_\tau = \left[B_\tau^{(1)} \quad \cdots \quad B_\tau^{(p)}\right] = JB_\tau$, and

$$\begin{aligned} a &= \left[\mu_1 \quad \cdots \quad \mu_q\right], & b &= \left[b_1 \quad \cdots \quad b_q\right], \\ \Phi &= P_{np} B, & \Psi &= \left(I_q \otimes J'\right) P_n\, \mu, \\ P_\tau &= \left(P \otimes I_\tau\right), & C_\tau &= \left(1_q{}' \otimes I_\tau\right), \end{aligned} \tag{23}$$

where $C_\tau$ is a $\tau \times \tau q$, matrix, $P_\tau$ is a $\tau q \times \tau q$ matrix, $\Phi$ is a $npq \times npq$ matrix containing parameters, $\Psi$ is a $npq \times q$ matrix, and $1_q$ denotes a $q \times 1$ vector of ones.

**Lemma 4.3**

Let the $npq \times 1$ vector $\tilde{Y}_t = \left(\xi_{t|t} \otimes Y_t\right)$. We have that

$$\mathrm{E}\left[y_{t+j} \middle| \mathcal{Y}_t\right] = \begin{cases} aP\xi_{t|t} + bP_{np}\tilde{Y}_t & \text{for } j = 1 \\ \left(aP^j + b\sum_{m=0}^{j-2} \Phi^m \Psi P^{j-1-m}\right)\xi_{t|t} + b\Phi^{j-1} P_{np} \tilde{Y}_t & \text{for } j \geq 2. \end{cases} \tag{24}$$

**Proof:** Substitute (19) and (20) into (21). The details are in the appendix. □ Note that we do not require $\Phi$ or $P$ to be non-singular[1].

This lemma is the building block of all results in the paper. In particular, it allows us to prove our main result, which gives a compact expression for the hypothesis in (10).

**Proposition 4.1**

For equation (10) to hold, the parameters of the VAR must satisfy the restrictions

$$\Xi^{(k)} P_{np} = 0, \text{ and } \Lambda^{(k)} - \lambda' = 0, \tag{25}$$

where $\Xi^{(k)} = \sum_{j=0}^{k} \Xi_j$ is $s \times npq$, $\Lambda^{(k)} = \sum_{j=1}^{k} \Lambda_j$ is $s \times q$, and

$$\Xi_j = \begin{cases} N_0 J\, C_{np} & \text{when } j = 0 \\ N_j b \Phi^{j-1} & \text{when } j \geq 1, \end{cases} \tag{26}$$

$$\Lambda_j = \begin{cases} N_1 aP & \text{when } j = 1 \\ N_j aP^j + N_j b \sum_{m=0}^{j-2} \Phi^m\, \Psi\, P^{j-1-m} & \text{when } j \geq 2. \end{cases} \tag{27}$$

**Proof:** Substitute (24) into (9). The details are in the appendix. □

## 4.1 Interpreting the Restrictions

---

[1] For a square matrix $X$, whether singular or not, we define $X^0 \equiv I$.

Let us focus on the restrictions on the autoregressive parameters given by $\Xi^{(k)}P_{np}=0$, as these are usually of more interest in hypothesis testing. There are $snpq$ equations, but there is no unique way in which these restrictions will be satisfied. Some combinations of the parameters contain exactly $snpq$ restrictions on the parameters, but it is possible to have fewer. This will be illustrated in the examples below. The maximum allowable number of restrictions for the Wald test that we can have on the autoregressive parameters is $n^2pq+q(q-1)$ (the number of parameters contained in $b$ and $P$).

One obvious way in which the restrictions can hold is if $\Xi^{(k)}=0$, giving a total of exactly $snpq$ restrictions on $b$. We will argue that this is the most interesting case, with some useful properties which will be examined below; another is if $\Xi^{(k)}$ and $P_{np}$ are orthogonal. For this to occur, $P$ must have reduced rank, which can give rise to fewer than $snpq$ restrictions.

There are of course a number of ways in which $P$ can have reduced rank, but the most straightforward case occurs when we have a serially uncorrelated Markov Chain (SUMC). For example, with $q=3$

$$P=\begin{pmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{32} \\ p_{13} & p_{23} & p_{33} \end{pmatrix}=\begin{pmatrix} p_{11} & p_{11} & p_{11} \\ p_{12} & p_{12} & p_{12} \\ p_{13} & p_{13} & p_{13} \end{pmatrix} \tag{28}$$

where $p_{\tau q}=1-p_{\tau 1}-p_{\tau 2}$. A simple way to obtain a SUMC is to impose $(q-1)^2$ restrictions on $P$ such that

$$P=\begin{pmatrix} p_1 & p_2 & \cdots & p_q \end{pmatrix}=\begin{pmatrix} p_\tau & p_\tau & \cdots & p_\tau \end{pmatrix}=1_q'\otimes p_\tau \tag{29}$$

where $p_\tau=\left[p_{\tau 1}\cdots p_{\tau q}\right]'$, $p_{\tau q}=1-\sum_{j=1}^{q-1}p_{\tau j}$ for some $\tau\in\{1,\cdots,q\}$.

To illustrate the use of proposition 4.1, let us consider a few examples. As a matter of notation, let $B_\tau^{(i,j,l)}$ be the $i,j{:}th$ element of the lag matrix $l$ in regime $\tau$. We consider the case when $n=q=2$, and $p=1$, whence

$$b=\left[\begin{array}{cc:cc} B_1^{(1,1)} & B_1^{(1,2)} & B_2^{(1,1)} & B_2^{(1,2)} \\ B_1^{(2,1)} & B_1^{(2,2)} & B_2^{(2,1)} & B_2^{(2,2)} \end{array}\right], \tag{30}$$

where we have dropped the (redundant) superscript for the lag matrix. In what follows, unless indicated otherwise, we confine the discussion to restrictions on the autoregressive parameters.

**Example 1 continued: term structure**

From (25), we see that parametric restrictions on the autoregressive parameters implied by the expectations hypothesis in example 1 are given by

$$e_1' C_2 - 0.5 e_2' b P_2 = 0, \tag{31}$$

since $C_{np} P_{np} = C_{np}$. The restrictions in (31) are

$$\begin{cases} p_{11} B_1^{(2,1)} + p_{12} B_2^{(2,1)} = 2 \\ p_{21} B_1^{(2,1)} + p_{22} B_2^{(2,1)} = 2 \\ p_{11} B_1^{(2,2)} + p_{12} B_2^{(2,2)} = 0 \\ p_{21} B_1^{(1,1)} + p_{22} B_2^{(2,2)} = 0. \end{cases} \tag{32}$$

We can write (32) as

$$\begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix} \begin{pmatrix} B_1^{(2,1)} \\ B_2^{(2,1)} \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \tag{33}$$

and

$$\begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix} \begin{pmatrix} B_1^{(2,2)} \\ B_2^{(2,2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{34}$$

since $p_{i1} + p_{i2} = 1$ (note that the matrix with the transition probabilities is the transpose of $P$). There are now two non-exclusive ways in which (33) and (34) can hold, either

$$\begin{cases} p_{11} + p_{22} = 1 \\ p_{11} B_1^{(2,1)} + (1-p_{11}) B_2^{(2,1)} = 2 \\ p_{11} B_1^{(2,2)} + (1-p_{11}) B_2^{(2,2)} = 0 \end{cases} \quad \text{or} \quad \begin{cases} B_1^{(2,1)} = 2 \\ B_2^{(2,1)} = 2 \\ B_1^{(2,2)} = 0 \\ B_2^{(2,2)} = 0. \end{cases} \tag{35}$$

The latter case with four linear restrictions ($snpq = 4$) corresponds to $\Xi^{(k)} = 0$, and requires that selected elements in the lag matrix are equal *across* regimes. The former represents a reduced rank condition on $P$, yielding three non-linear restrictions. In the case of two regimes, this reduced rank condition is the same as requiring the MC to be serially uncorrelated. Note that with a one period forecast horizon ($k = 1$), $\Xi^{(1)}$ does not contain $P$, and so the two alternatives are non-exclusive. For $k > 1$ this will not be the case.

For the restrictions to hold when there are more than two regimes, the cases with rank one and full rank for $P$ remain, but there are some intermediate possibilities as well. This is best illustrated by extending the above example to three regimes. The corresponding expression to (32) with three regimes is

$$\begin{cases} p_{11}B_1^{(2,1)} + p_{12}B_2^{(2,1)} + (1-p_{11}-p_{12})B_3^{(2,1)} = 2 \\ p_{21}B_1^{(2,1)} + p_{22}B_2^{(2,1)} + (1-p_{21}-p_{22})B_3^{(2,1)} = 2 \\ p_{31}B_1^{(2,1)} + p_{32}B_2^{(2,1)} + (1-p_{31}-p_{32})B_3^{(2,1)} = 2 \\ p_{11}B_1^{(2,2)} + p_{12}B_2^{(2,2)} + (1-p_{11}-p_{12})B_3^{(2,2)} = 0 \\ p_{21}B_1^{(2,2)} + p_{22}B_2^{(2,2)} + (1-p_{21}-p_{22})B_3^{(2,2)} = 0 \\ p_{31}B_1^{(2,2)} + p_{32}B_2^{(2,2)} + (1-p_{31}-p_{32})B_3^{(2,2)} = 0. \end{cases} \tag{36}$$

Clearly, one way for these restrictions to hold is if $B_j^{(2,1)} = 2$ and $B_j^{(2,2)} = 0$ for $j = \{1,2,3\}$. This corresponds to $\Xi^{(k)} = 0$, which yields a total of six restrictions $(snpq = 6)$.

A SUMC is obtained if $p_{11} = p_{21} = p_{31}$, $p_{12} = p_{22} = p_{32}$ , giving four restrictions on *P*, and

$$\begin{cases} p_{11}B_1^{(2,1)} + p_{12}B_2^{(2,1)} + (1-p_{11}-p_{12})B_3^{(2,1)} = 2 \\ p_{11}B_1^{(2,2)} + p_{12}B_2^{(2,2)} + (1-p_{11}-p_{12})B_3^{(2,2)} = 0, \end{cases} \tag{37}$$

which also gives a total of six restrictions (this is just a coincidence).

Let us consider the intermediate case when *P* has rank two. One way in which this can occur is if $p_{11} = p_{21} \neq p_{31}$ , $p_{12} = p_{22} \neq p_{32}$, giving the four non-redundant restrictions

$$\begin{pmatrix} p_{11} & p_{12} & 1-p_{11}-p_{12} \\ p_{31} & p_{32} & 1-p_{31}-p_{32} \end{pmatrix} \begin{pmatrix} B_1^{(2,1)} \\ B_2^{(2,1)} \\ B_3^{(2,1)} \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \tag{38}$$

$$\begin{pmatrix} p_{11} & p_{12} & 1-p_{11}-p_{12} \\ p_{31} & p_{32} & 1-p_{31}-p_{32} \end{pmatrix} \begin{pmatrix} B_1^{(2,2)} \\ B_2^{(2,2)} \\ B_3^{(2,2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{39}$$

Again we obtain 6 restrictions on the parameters.

This example serves to illustrate two important points. First, having *snpq* restrictions on the parameters is not the only possibility: there can be fewer, since not all restrictions need be independent when $\Xi^{(k)}$ and $P_{np}$ are orthogonal. Second, to obtain a SUMC in the above example (for given $n$ and $p$) does not require more than $snpq$ restrictions, but this can occur for larger number of regimes. For example, with $q = 4$, we would obtain 8 restrictions corresponding to $\Xi^{(k)} = 0$, and 11 for the SUMC case. There are then 3 overidentifying restrictions on the parameters above that implied by RE. With more regimes the number of overidentifying restrictions increases rapidly, a feature due to requiring $(q-1)^2$ restrictions to impose a SUMC. Consequently, only when we have a small number of regimes is it possible to impose SUMC without imposing restrictions unwarranted by RE, though we can still consider the "intermediate" rank cases for $P$.

**4.2 Long Forecast Horizon**

In the example above, the short horizon makes the problem manageable, but for many applications proposition 4.1 is too general to be of direct use. We need to narrow the class of $N_j$ considered, and obtain simpler expressions. One convenient way to do this is to let $N_j = -\delta^j N$ for $j \geq 1$, where $N$ is a $s \times n$ matrix of constants and $\delta \in (0,1)$ is a discount factor. Formally, we focus on hypotheses of the type

$$N_0\, y_t - N \sum_{j=1}^{k} \delta^j \mathrm{E}\left[ y_{t+j} \middle| \mathcal{Y}_t \right] = 0, \tag{40}$$

where in many applications $N = e_i'$, including the ones considered below, but we do not need to make such an assumption to get useful results.

The next corollary shows what happens to the restrictions when we consider hypotheses spanning a long horizon, which allows us, for instance, to consider perpetuity models. The practical use of corollary 4.1 is to provide expressions for the restrictions that only require matrix multiplication (i.e. without summation signs).

**Corollary 4.1**

Let $N_j = -\delta^j N$, $\delta \in (0,1)$, and $\overline{\Phi}_\tau = I_{npq} - (\delta\,\Phi)^\tau$. If $\overline{\Phi}_1$ is non-singular, then

$$\Xi^{(k)} = N_0 J C_{np} - N b \delta\, \overline{\Phi}_k \overline{\Phi}_1^{-1}, \tag{41}$$

**Proof:** $\Xi^{(k)} = N_0 J C_{np} - S$ where $S \equiv N b \delta \left[ I_{npq} + \delta\,\Phi + \cdots + (\delta\,\Phi)^{k-1} \right]$. Solving for $S$, we find $S - S\delta\,\Phi = N b \delta \left( I_{npq} - (\delta\,\Phi)^k \right)$. Thus, $S = N b \delta \left( I_{npq} - (\delta\,\Phi)^k \right) \left( I_{npq} - \delta\,\Phi \right)^{-1}$, and the result follows. □

This next corollary to proposition 4.1 considers the special case of perpetuity models, which is of interest for a large number of hypotheses.

**Corollary 4.2**

Let the $snpq \times n^2 pq$ matrix $R = I_{npq} \otimes \delta\left(N_0 + N\right)$, the $snpq \times 1$ vector $r = \mathrm{vec}\, N_0 J C_{np}$, and $\Xi = \Xi^{(\infty)}$. Assume that the largest eigenvalue of $\delta\,\Phi$ is inside the unit circle, then when $k \to \infty$ the restrictions on the autoregressive parameters in (25) hold if

$$R\, \mathrm{vec}\, b = r, \tag{42}$$

**Proof:** Use (41) to obtain $\Xi P_{np} = \left[N_0 JC_{np} - Nb\delta\overline{\Phi}_1^{-1}\right]P_{np} = 0$. This holds if $N_0 JC_{np} - Nb\delta\overline{\Phi}_1^{-1} = 0$. Post-multiply by $\overline{\Phi}_1$: $\delta N_0 JC_{np}\Phi + Nb\delta - N_0 JC_{np} = 0$. Now, $JC_{np}\Phi = b$ and thus $\delta\left(N_0 + N\right)b - N_0 JC_{np} = 0$. □

This corollary has two remarkable implications. First, the restrictions are linear, and second, they do not involve the transition probabilities. Both of these are a bit surprising since the $j{:}th$ term in the summation is $Nb\delta(\delta P_{np}B)^{j-1}$, which involves the transition probabilities. What the corollary tells us is that these probabilities "wash out" when we let $k \to \infty$. This result can be seen as an extension of the test on the autoregressive parameters in Campbell & Shiller (1987) with a linear VAR, in the sense that if we let $q = 1$ we obtain exactly their form of linear restrictions for the present value model of the term structure.

**Example 2 continued: asset pricing**

Recall that the hypothesis in example 2 was $e_1' y_t - \delta(1-\delta)^{-1} e_2' \sum_{j=1}^{\infty} \delta^j \mathrm{E}\left[y_{t+j} \middle| \mathcal{Y}_t\right] = 0$. In the notation of this section, we have $N_0 = e_1'$ and $N = \delta^* e_2'$, where $\delta^* = \delta(1-\delta)^{-1}$, which implies that $r = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix}'$, and $R = \begin{bmatrix}\delta & \delta^*\end{bmatrix}$. Using (42), the restrictions on the autoregressive parameters implied by $\Xi = 0$ are then

$$\begin{cases} \delta B_1^{(1,1)} + \delta^* B_1^{(2,1)} = 1 \\ \delta B_2^{(1,1)} + \delta^* B_2^{(2,1)} = 1 \\ \delta B_1^{(1,2)} + \delta^* B_1^{(2,2)} = 0 \\ \delta B_2^{(1,2)} + \delta^* B_2^{(2,2)} = 0. \end{cases} \tag{43}$$

Note that the form of the restrictions here are very different to the example with the expectations hypothesis. In (35), the restriction $\Xi^{(k)} = 0$ required selected elements of the lag matrix to be equal *across* regimes. In (43), by contrast, the restrictions are on selected elements of the lag matrix *within* each regime.

## 5 State Dependent Discount Factor

The hypotheses embodied in (9) are fairly general, but we may want to let the $N_j$ terms be state dependent. In particular, the case when the discount factor depends on the state might be useful. There are of course many other ways in which an individual's time preference can change over time, but a discount factor that depends on the unobservable state might provide a approximation. Having more than one exogenous discount factor to change might provide more information about the model's performance. In particular, if a very large/small discount factor is needed to not reject some hypothesis, then that might be construed as further evidence against it.

Let $N_j = -N\prod_{\tau=1}^{j}\delta_{s_{t+\tau}}$. One way to extend (40) is then

$$N_0\, y_t - N\sum_{j=1}^{k}\mathrm{E}\left[\left(\prod_{\tau=1}^{j}\delta_{s_{t+\tau}}\right)y_{t+j}\middle|\mathcal{Y}_t\right] = 0. \tag{44}$$

The $j{:}th$ term in the summation for $j \geq 2$ is given by

$$\begin{aligned}
\mathrm{E}\left[\left(\prod_{\tau=1}^{j}\delta_{s_{t+\tau}}\right)y_{t+j}\middle|\mathcal{Y}_t\right] &= \sum_{i_j=1}^{q}\cdots\sum_{i_1=1}^{q}\left(\prod_{\tau=1}^{j}\delta_{i_\tau}\right)\mathrm{E}\left[y_{t+j}\middle|s_{t+1}=i_1,\cdots,s_{t+j}=i_j,\mathcal{Y}_t\right]\times \\
&\qquad \mathrm{pr}\left[s_{t+j}=i_j,\cdots,s_{t+1}=i_1\middle|\mathcal{Y}_t\right] \\
&= \sum_{i_j=1}^{q}\cdots\sum_{i_1=1}^{q}\left(\delta_{i_1}\cdots\delta_{i_j}\right)\left[\mu_{i_j} + J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h}B_{j+1-m}\right)J'\,\mu_{j-h} + J\left(\prod_{m=1}^{j}B_{j+1-m}\right)Y_t\right]\times \\
&\qquad \left(\prod_{f=0}^{j-2}(e_{i_{j-f}}{}'\,Pe_{i_{j-f-1}})(e_{i_1}{}'\,P\xi_{t|t})\right).
\end{aligned} \tag{45}$$

Let

$$\delta = \mathrm{diag}\left(\delta_1,\cdots,\delta_q\right), \quad \text{and} \quad \delta_\tau = \delta \otimes I_\tau. \tag{46}$$

**Proposition 5.1**

For $j \geq 1$, we can replace (26) by

$$\Xi_j = Nb\delta_{np}(\Phi\delta_{np})^{j-1}. \tag{47}$$

**Proof:** Similar to the proof of proposition 4.1 and omitted. □

**Corollary 5.1**

Let $R_\delta = \delta_{np}\otimes\left(N_0 + N\right)$, and $\Xi_j$ be defined from (47), and assume that the largest eigenvalue of $\Phi\delta_{np}$ is inside the unit circle. Then the parametric restrictions in $\Xi = 0$ can be written as

$$R_\delta \mathrm{vec}\, b = r. \tag{48}$$

**Proof:** Similar to proof of corollary 4.2 and omitted. □

Note that although the discount factor is state dependent, we still get linear restrictions on the autoregressive parameters.

# 6 Testing the Restrictions

## 6.1. The Wald test

In general, the restrictions on the parameters derived in the previous section are non-linear, and involve the Markov transition probabilities. Estimation of the VAR subject to these restrictions is likely to be cumbersome, and thus when we turn to testing them the Wald framework seems most appropriate, since it entails estimating only the unconstrained model. The Wald test, however, is not invariant to transformations of the restrictions in finite samples, see for example Gregory and Veall (1985). The non-linear restrictions we consider are subject to this criticism, and so whenever possible it is desirable to estimate the constrained model as well. We could then compute Lagrange Multiplier (LM) or Loglikelihood Ratio (LR) tests for comparison with the Wald.

The object of this section is to consider statistical tests of the restrictions in (25). To be concrete, suppose

$$T^{1/2}(\hat{\theta}-\theta_0)\xrightarrow{\mathrm{d}}\mathrm{N}_d(0,\mathrm{Q}), \tag{49}$$

where $\theta_0$ $(d\times 1)$ is the vector 'true' values of the population parameters for the regime shift VAR($p$), $\hat{\theta}$ is its maximum likelihood estimate (MLE), and Q is the covariance matrix.

Let $g(\theta_0)=0$ be a $r\times 1$ vector of restrictions on $\theta$ and let the $r\times d$ matrix

$$G=\frac{\partial g}{\partial \theta'}. \tag{50}$$

The Wald test is then given by

$$\mathrm{W}=g(\hat{\theta})'\left[\hat{G}\hat{\mathrm{Q}}\hat{G}'\right]^{-1}g(\hat{\theta})\xrightarrow{\mathrm{d}}\chi^2(r)\ \text{on}\ \mathrm{H}_0\text{:}\ \ g(\theta_0)=0, \tag{51}$$

where $\hat{Q}$ is an estimate of the covariance matrix $Q$, and $\hat{G}=G(\hat{\theta})$. Clearly, a necessary condition for the test to be well defined is that $r\leq d$, i.e. there cannot be more restrictions than parameters. Moreover, $G(\hat{\theta})$ must have full row rank with probability one. If either of these conditions fail $\hat{G}\hat{\mathrm{Q}}\hat{G}'$ will be singular in the limit and the asymptotic distribution of W unknown. Further, as a regularity condition we require that $G$ not change rank in a neighbourhood of $\hat{\theta}$.

Calculation of $\hat{\theta}$ is done with the methods described in Lindgren (1978), Holst, Lindgren, Holst, and Thuvesholmen (1994), and Hamilton (1990,1994), and will not be dwelt on further here. Consistency of the MLE $\hat{\theta}$ in a VAR with Markov switching has only been shown quite recently in a paper by Krishnamurthy and Rydén (1996). Asymptotic normality of the MLE is still an open question, but we will nevertheless assume it; Bickel, Ritov, and Rydén (1997) have proved asymptotic normality of the MLE for a Markov switching model with no autoregression.

To implement the test above we need the first derivatives of the constraints with respect to the parameters in the model. For the restrictions on the autoregressive parameters, the relevant parameter vector is

$$\theta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \operatorname{vec} b \\ \operatorname{vec} \Gamma \end{bmatrix}, \tag{52}$$

which is of dimension $d = n^2 pq + q(q-1)$, and where the $q-1 \times q$ matrix

$$\Gamma = \begin{pmatrix} p_{11} & \cdots & p_{q1} \\ \vdots & & \vdots \\ p_{1q-1} & \cdots & p_{qq-1} \end{pmatrix} \tag{53}$$

contains only the unique elements of $P$.

What is an appropriate way to formulate the Wald statistic? It was shown above that if $\Xi^{(k)} = 0$, the RE restrictions are satisfied, but it was also demonstrated that this was not the only way they might hold. In particular, with reduced rank on $P$ the restrictions may also hold if $\Xi^{(k)}$ and $P$ are orthogonal. As illustrated in the examples, the number of restrictions being tested under the null depends crucially on which case is being considered. If we were to test $\Xi^{(k)} P_{np} = 0$ the test would have "different degrees of freedom" depending on if the matrices were orthogonal or not, and such a statistic is not well defined.

There seems to be at least two good reasons for focusing on testing $\Xi^{(k)} = 0$ only. First, it seems that most of the MC estimated in the literature are serially correlated, such as those in Hamilton (1988,1989), Engel and Hamilton (1990), Hamilton and Susmel (1994). On the strength of this it seems innocuous to assume that the MC is serially correlated as part of the maintained hypothesis (this can anyway be tested). Further, it is clear that when $P$ has full rank, $\Xi^{(k)} = 0$ is the *only* way the RE restrictions can be satisfied. Second, we gain much in terms of simplicity. This is most apparent when considering the infinite horizon case, where RE restrictions are linear.

Nevertheless, a SUMC can sometimes have interesting economic implications. For example, the conditional term premium in Blix (1996) for a correlated MC is conditionally heteroskedastic; for a SUMC, by contrast, it is white noise. However, if we want to impose the additional restrictions of a SUMC, we cannot just append it to $\text{vec}\left(\Xi^{(k)}P_{np}\right)$, because not all the restrictions are independent. In the example above with $n=2$, $p=1$, and $q=2$ we had 4 restrictions when testing $\Xi^{(k)}=0$, but 3 when also imposing a SUMC. What this means is that the restrictions of a SUMC has to be substituted into $\text{vec}\left(\Xi^{(k)}P_{np}\right)$ to obtain only the independent restrictions[2]. Although this is easy for the example above, in general it might be quite difficult.

Let us now turn to formally introducing the RE hypothesis we want to test. In the rest of the paper, we will focus on tests of $\text{vec}\Xi^{(k)}=0$. Implicit in the rest if the discussion is also that for hypotheses spanning a finite horizon, different weights $N_j$ are used, while for the infinite horizon case, a constant declining weight is assumed.

Formally, we can test the hypothesis that

$$\text{H}_0^{\text{A}}: \quad \Xi^{(k)}=0. \tag{54}$$

in which case, with the assumptions in corollary 4.2., we have

$$g^{\text{A}}(\theta)=\begin{cases}\text{vec}\,\Xi^{(k)} & k<\infty\\ R\beta-r & k=\infty.\end{cases} \tag{55}$$

A necessary requirement for the test is that $snpq<n^2pq+q(q-1)$, i.e. the number of restrictions should be less than the number of parameters.

Finally, let us briefly discuss how an hypothesis that includes the restrictions on the intercept. In this case the $qn+n^2pq+q(q-1)$ vector with parameters takes the form

$$\theta=\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix}, \tag{56}$$

where $\alpha=\text{vec}\,a$. The complete set of RE restrictions are then

---

[2] If this is not done $G$, the first derivative of $g$, will not have full row rank, and the Wald test will not be well defined.

$$\mathrm{H}_0^{\mathrm{B}}: \quad \Xi^{(k)} = 0 \quad \text{and} \quad \Lambda^{(k)} - \lambda' = 0, \tag{57}$$

so that

$$g^{\mathrm{B}}(\theta) = \begin{bmatrix} g^{\mathrm{A}}(\theta) \\ \mathrm{vec}\left[\Lambda^{(k)} - \lambda'\right] \end{bmatrix}. \tag{58}$$

We now require that $snpq + sq < qn + n^2 pq + q(q-1)$.

**6.2. Finding the derivatives of *g***

The rest of this section is devoted to introducing the notation needed to find the first derivative of $g(\theta)$, and can be skipped without loss of continuity. At the end of this section we display different versions of $G$, based on the hypotheses discussed above, and also on the horizon.

First, let the $q^2 \times q(q-1)$ matrix $\Upsilon$ be such that $\mathrm{vec}\,\mathrm{d}P = \Upsilon\,\mathrm{d}\gamma$, for which an explicit expression is $\Upsilon = I_q \otimes \begin{pmatrix} I_{q-1} & -1_{q-1} \end{pmatrix}'$. Further, let the $k_{m,n}$ $(mn \times mn)$ commutation matrix be defined by the property that $k_{m,n}\,\mathrm{vec}\,\mathrm{X} = \mathrm{vec}\,\mathrm{X}'$ for any $m \times n$ matrix $\mathrm{X}$ (when $\mathrm{X}$ is square we simply write $k_n$). An explicit expression for $k_{m,n}$ is $k_{m,n} = \sum_{i=1}^{m}\left(e_i \otimes I_n \otimes e_i'\right)$, see Magnus (1988, p 38). Finally, define the $(\tau q)^2 \times q^2$ matrix $\kappa_\tau \equiv (I_q \otimes k_{\tau,q} \otimes I_\tau)(I_{q^2} \otimes \mathrm{vec}\, I_\tau)$, and the $rsq^2 \times rsq$ matrix $\Delta_{r,s} \equiv \left(I_q \otimes k_{s,q} \otimes I_r\right)\sum_{i=1}^{q}\left((\mathrm{vec}\, E_{ii})e_i' \otimes I_{rs}\right)$. When $r = s$, we simply write $\Delta_r$. We can now give the required derivatives for the above hypotheses.

**Proposition 6.1**

Consider testing the hypothesis in (54) that $\Xi^{(k)} = 0$. When $k$ is infinite the matrix with first derivatives takes the simple form

$$G^{\mathrm{A}} = \begin{bmatrix} R & 0 \end{bmatrix} \tag{59}$$

while when $k$ is finite we have that

$$G_k^{\mathrm{A}} = \sum_{j=0}^{k}\begin{pmatrix} \dfrac{\partial\,\mathrm{vec}\,\Xi_j}{\partial\beta'} & \dfrac{\partial\,\mathrm{vec}\,\Xi_j}{\partial\gamma'} \end{pmatrix}, \tag{60}$$

where for $j \geq 2$,

$$\begin{cases} \dfrac{\partial\,\mathrm{vec}\,\Xi_j}{\partial\beta'} = \left((\Phi^{j-1})' \otimes N_j\right) + \sum_{m=1}^{j-1}\left((\Phi^{j-1-m})' \otimes N_j b \Phi^{m-1} P_{np}\right)\Delta_{np}\left(I_{npq} \otimes J'\right) \\ \dfrac{\partial\,\mathrm{vec}\,\Xi_j}{\partial\gamma'} = \sum_{m=1}^{j-1}\left((B\Phi^{j-1-m})' \otimes N_j b \Phi^{m-1}\right)\kappa_{np}\Upsilon, \end{cases} \tag{61}$$

while for $j = \{0,1\}$,

$$\begin{cases} \dfrac{\partial \operatorname{vec} \Xi_1}{\partial \beta'} = I_{npq} \otimes N_1 \\ \dfrac{\partial \operatorname{vec} \Xi_1}{\partial \gamma'} = 0 \end{cases} \qquad \begin{cases} \dfrac{\partial \operatorname{vec} \Xi_0}{\partial \beta'} = 0 \\ \dfrac{\partial \operatorname{vec} \Xi_0}{\partial \gamma'} = 0. \end{cases} \tag{62}$$

**Proof:** In the appendix. □

**Proposition 6.2**

Consider testing the hypothesis in (57). The matrix with first derivatives of (58) is

$$G = \sum_{j=0}^{k} \begin{bmatrix} 0 & \dfrac{\partial \operatorname{vec} \Xi_j}{\partial \beta'} & \dfrac{\partial \operatorname{vec} \Xi_j}{\partial \gamma'} \\ \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \alpha'} & \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \beta'} & \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \gamma'} \end{bmatrix}, \tag{63}$$

where $\Lambda_0 = 0$. The derivatives in the top submatrix are already given above, and the other required derivatives are, for $j \geq 3$,

$$\begin{cases} \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \alpha'} = \left((P^j)' \otimes N_j\right) + \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b\, \Phi^m\right)\left(I_q \otimes P \otimes J'\right)\Delta_{n,1} \\ \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \beta'} = \sum_{m=0}^{j-2} \left((\Phi^m \Psi P^{j-1-m})' \otimes N_j\right) \\ \qquad + \sum_{m=1}^{j-2} \sum_{s=1}^{m} \left((\Phi^{m-s} \Psi P^{j-1-m})' \otimes N_j b \Phi^{s-1} P_{np}\right) \Delta_{np} \left(I_{npq} \otimes J'\right) \\ \dfrac{\partial \operatorname{vec} \Lambda_j}{\partial \gamma'} = \sum_{m=1}^{j} \left((P^{j-m})' \otimes N_j a\, P^{m-1}\right)\Upsilon + \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right)\left(\mu' \otimes I_q \otimes J'\right)\kappa_n \Upsilon \\ \qquad + \sum_{m=0}^{j-2} \sum_{s=1}^{j-1-m} \left((P^{j-1-m-s})' \otimes N_j b\, \Phi^m \Psi P^{s-1}\right)\Upsilon \\ \qquad + \sum_{m=1}^{j-2} \sum_{s=1}^{m} \left((B\Phi^{m-s} \Psi P^{j-1-m})' \otimes N_j b \Phi^{s-1}\right)\kappa_{np} \Upsilon, \end{cases} \tag{64}$$

while for $j = \{1,2\}$,

$$\begin{cases} \dfrac{\partial \operatorname{vec} \Lambda_2}{\partial \alpha'} = \left((P^2)' \otimes N_2\right) + \left(P' \otimes N_2 b\right)\left(I_q \otimes P \otimes J'\right)\Delta_{n,1} \\ \dfrac{\partial \operatorname{vec} \Lambda_2}{\partial \beta'} = \left((\Psi P)' \otimes N_2\right) \\ \dfrac{\partial \operatorname{vec} \Lambda_2}{\partial \gamma'} = \sum_{m=1}^{2} \left((P^{2-m})' \otimes N_2 a P^{m-1}\right)\Upsilon + \left(I_q \otimes N_2 b \Psi\right)\Upsilon \\ \qquad + \left(P' \otimes N_2 b\right)\left(\mu' \otimes I_q \otimes J'\right)\kappa_n \Upsilon \end{cases} \qquad \begin{cases} \dfrac{\partial \operatorname{vec} \Lambda_1}{\partial \alpha'} = \left(P' \otimes N_1\right) \\ \dfrac{\partial \operatorname{vec} \Lambda_1}{\partial \beta'} = 0 \\ \dfrac{\partial \operatorname{vec} \Lambda_1}{\partial \gamma'} = \left(I_q \otimes N_1 a\right)\Upsilon. \end{cases} \tag{65}$$

**Proof:** In the appendix. □

## 7 Expectations Hypothesis

In this section the methods discussed above will be applied to testing the expectations hypothesis of the term structure. We test the version of the EH discussed in example 1 above, which states that the yield on a two period discount bond should equal the average expected return of holding two one-period discount bonds. This test has been performed in a paper by Sola & Driffill (1994) for US treasury bills also in a model with regime shifts.

As noted above, the form of *centered* regime shift model used in their paper differs from the one considered here. What does this imply for comparing results in general? Since we are testing the autoregressive parameters, this does not make any difference for the *form* of the restrictions, i.e. the tests look identical although the models are different (as discussed at the end of section 2). The *estimated* dynamics, however, will be different for the two models, which will of course affect the test statistics.

Unfortunately, this does not mean that we can directly compare the restrictions, because the model in Sola & Driffill can be seen as a special case of (8) where the $i{:}th$ lag is replaced by a state invariant matrix, i.e. $B^{(i)}_{s_{t-i}} = B^{(i)} \quad \forall i \in \{1, \cdots, p\}$, which, of course, has more restrictive dynamics. The results in this paper could still be used to test such restrictions, but then they are identical in form to those in linear VAR:s - even if the drift and the covariance terms are state dependent.

This gives rise to an important difference in the form of the restrictions. The Sola & Driffill restrictions on the autoregressive parameters are identical to those in standard VAR:s, while here all the restrictions on the autoregressive parameters are *across* states. The only restriction on regime dependent terms in Sola & Driffill is on the drift term, which might be interpreted as requiring the premium to be equal across states. A constant premium in our setting would require

$$\mu_{21} p_{11} + \mu_{22}(1 - p_{11}) = \mu_{21}(1 - p_{22}) + \mu_{22} p_{22}, \qquad (66)$$

where $\mu_{i\tau}$ is the drift term for the $i{:}th$ equation in state $\tau$. This restrictions is satisfied if either $\mu_{21} = \mu_{22}$ or $p_{11} = 1 - p_{22}$, interpreted as requiring that selected drift terms should be equal across states and a SUMC respectively. Though such restrictions are of interest, they are not necessarily implied by the EH. Tests of the EH in the literature have focused on restrictions on the autoregressive parameters, and left the drift term unrestricted; see the discussion in Blix (1996).

The restrictions are different in another sense as well. Although the EH in this paper involves a one period ahead forecast, the Sola & Driffill procedure is to take expectations conditional on information at time $t-1$ to control for an assumed measurement error. This implies that the forecast of $\Delta R_{t+1}^{(1)}$ will be *two* periods ahead rather than one, which gives rise to non-linear restrictions. In this paper, the possibility of measurement error is not addressed, although the results in section 4 suggest how it might be done, e.g. condition both sides of (10) on $\mathcal{Y}_{t-1}$ instead. The advantage of ignoring it is that for the one period forecast horizon, the restrictions are linear; for all other finite horizons, they are non-linear.

We use the same frequency of data as Sola & Driffill, namely 3 and 6 month US bills on a quarterly basis; the data here is London Interbank Offert Rates (LIBOR) taken from the IMF for the period 63Q1-96Q3, and is plotted in levels on an annual yields basis in figure 1. The transformed data in terms of spread and first difference in the short rate are plotted in figure 2.

Figure 1. US Interest Rates

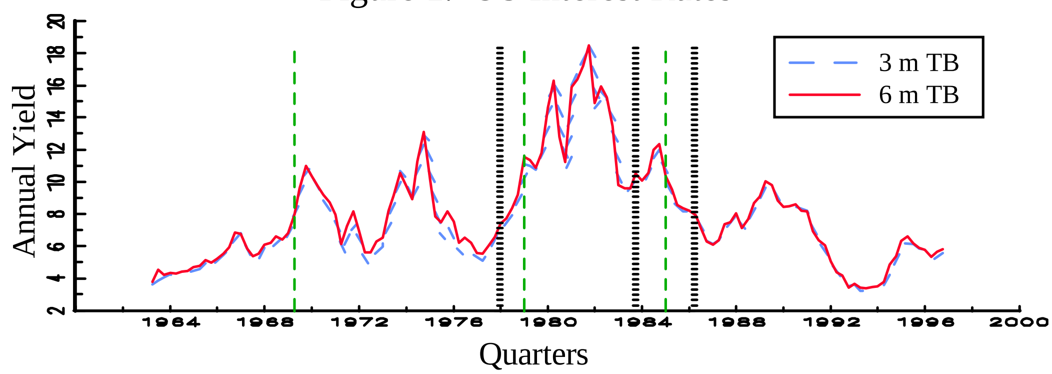
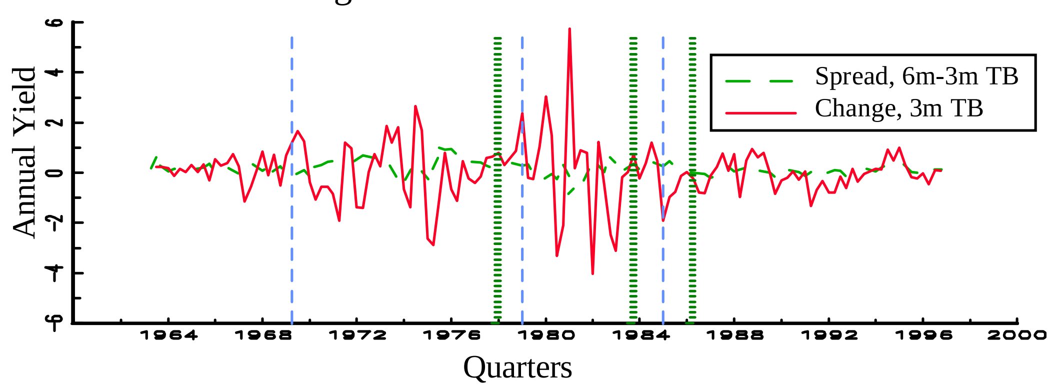



Figure 2. US Interest Rates



The next step is to estimate the regime shift VAR($p$) on $y_t = \begin{pmatrix} S_t & \Delta R_t^{(1)} \end{pmatrix}'$ (using exactly the same notation as in example 1). We choose two regimes, not because it is the only specification that can be justified, but because it is the simplest one that is of interest. Moreover, none of the diagnostics for the VAR with one lag presented below reject this specification at conventional levels.

**Table 1**

Model 1. State dependence in $\mu_{s_t}, B_{s_t}$, and $\Omega_{s_t}$.

$$L_1(\hat{\theta}) = -121.9$$

State 1

$$\begin{pmatrix} S_t \\ \Delta R_t^{(1)} \end{pmatrix} = \begin{pmatrix} \underset{(0.05)}{0.10} \\ \underset{(0.36)}{-0.16} \end{pmatrix} + \begin{pmatrix} \underset{(0.15)}{0.59} & \underset{(0.03)}{-0.04} \\ \underset{(1.10)}{0.41} & \underset{(0.20)}{0.16} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta R_{t-1}^{(1)} \end{pmatrix} \qquad \Omega_1 = \begin{pmatrix} \underset{(0.015)}{0.078} & \underset{(0.062)}{-0.231} \\ \underset{(0.062)}{-0.231} & \underset{(0.804)}{2.920} \end{pmatrix}$$

State 2

$$\begin{pmatrix} S_t \\ \Delta R_t^{(1)} \end{pmatrix} = \begin{pmatrix} \underset{(0.02)}{0.09} \\ \underset{(0.07)}{-0.16} \end{pmatrix} + \begin{pmatrix} \underset{(0.09)}{0.39} & \underset{(0.03)}{0.03} \\ \underset{(0.37)}{1.44} & \underset{(0.11)}{0.19} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta R_{t-1}^{(1)} \end{pmatrix} \qquad \Omega_2 = \begin{pmatrix} \underset{(0.004)}{0.015} & \underset{(0.007)}{0.014} \\ \underset{(0.007)}{0.014} & \underset{(0.042)}{0.208} \end{pmatrix}$$

$$P = \begin{pmatrix} \underset{(0.05)}{0.94} & \underset{(0.03)}{0.04} \\ \underset{(0.05)}{0.06} & \underset{(0.03)}{0.96} \end{pmatrix},$$

Wald test of Serially uncorrelated MC

W(4)=556 (0%)

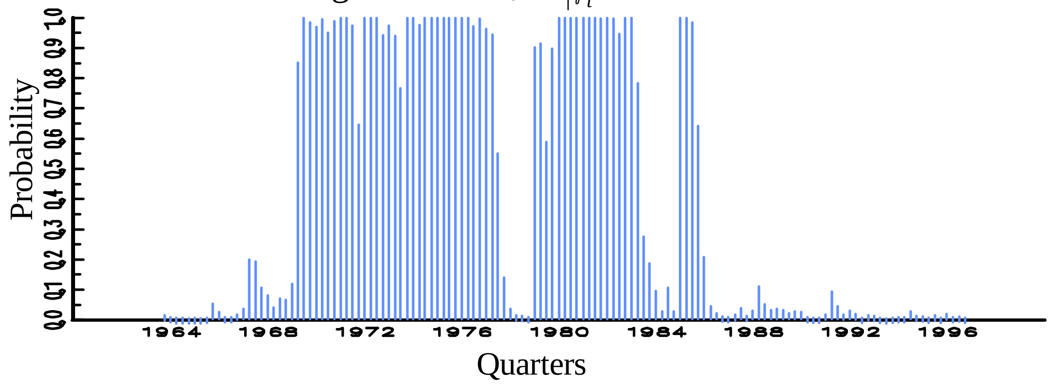
Misspecification Tests

| | Equation 1 | Equation 2 |
|---|---|---|
| Autocorrelation | F(4,122)=1.5 (20%), | F(4,122)=0.8 (51%) |
| ARCH | F(4,122)=1.4 (23%), | F(4,122)=2.4 (6%) |
| Markov | F(4,122)=1.3 (27%), | F(4,122)=1.5 (20%) |

The loglikelihood of the unrestricted estimates for model $\tau$ is denoted by $L_\tau(\hat{\theta})$; when the EH restrictions are imposed, $L_\tau(\tilde{\theta})$. The results for model 1 are displayed in table 1. Below each estimate, a standard error based on conditional scores is given in parentheses. Also displayed are misspecification tests for residual autocorrelation, ARCH effects, and for the Markov assumption. These are *F* versions of tests based on conditional scores from Newey (1985), Tauchen (1985), White (1987), and suggested by Hamilton (1996) for a univariate model. The significance level is given in parentheses.

In state 1, the shocks are negatively correlated and have a higher variance than in state 2. In state 2, the shocks are almost uncorrelated, and $S_{t-1}$ has much stronger effect on $\Delta R_t^{(1)}$. The probability of being in a particular state, $\text{pr}\left[s_t = 1 \middle| \mathcal{Y}_t\right]$, is plotted in figure 3, from which it can be seen that state 1 (with "high variance") covers most of the Volcker money growth target, as well as most of the 1970:s. Since the mid 1980:s, however, the observations all seem to belong to state 2.

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




These probabilities are also used to provide another type of information about when we switch from one regime to the other. A "switch" to a given state in time $t+1$ is defined as having occurred when the probability of being in that state is less than one-half at time $t$ , but more than one-half at time $t+1$. Each such event is depicted as a vertical line in figures 1,2, 4 and 5. The long dashes denote a switch into state 1, the short into state 2.

Table 2 displays a Wald test of the EH in (35) given by $\text{vec}\Xi^{(1)} = 0$ using variances based on conditional scores. Since the restrictions in this example are linear, it was not difficult to estimate the constrained model as well, allowing us to perform a LR test for restrictions for ready comparison to the Wald test: the likelihood value when imposing the EH on the model in table 1 is $L_1(\tilde{\theta}) = -128.8$. The evidence so far is ambiguous. On the one hand, based on the Wald test we do not reject the null at conventional levels, while for the LR we do. By contrast, the LR test for ten restrictions in Sola & Driffill could not be rejected at the 5% level (there are 4 lags in their VAR).

**Table 2**

Test of EH: $B_1^{(2,1)} = B_2^{(2,1)} = 2$ and $B_1^{(2,2)} = B_2^{(2,2)} = 0$.

W(4)=7.6 (11%)
LR(4)=2(128.8-121.9)= 13.8 (0.8%).

A remarkable feature of the estimates in table 1 is that all parameters in the drift and autoregressive terms except $B_i^{(2,1)}$ are virtually identical, suggesting that perhaps most of the models' improvement over the standard VAR comes from having a state dependent covariance matrix. To investigate this possibility, we estimated a VAR with state invariant drift and autoregressive parameters, displayed in table 3.

**Table 3**

Model 2, state dependent $\Omega_{s_t}$ only.

$$L_2(\hat{\theta}) = -125.0$$

$$\begin{pmatrix} S_t \\ \Delta R_t^{(1)} \end{pmatrix} = \begin{pmatrix} \underset{(0.02)}{0.07} \\ \underset{(0.08)}{-0.10} \end{pmatrix} + \begin{pmatrix} \underset{(0.06)}{0.57} & \underset{(0.02)}{-0.02} \\ \underset{(0.30)}{0.90} & \underset{(0.08)}{0.19} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta R_{t-1}^{(1)} \end{pmatrix}, \qquad P = \begin{pmatrix} \underset{(0.04)}{0.94} & \underset{(0.03)}{0.04} \\ \underset{(0.04)}{0.06} & \underset{(0.03)}{0.96} \end{pmatrix}$$

$$\Omega_1 = \begin{pmatrix} \underset{(0.015)}{0.085} & \underset{(0.052)}{-0.251} \\ \underset{(0.052)}{-0.251} & \underset{(0.616)}{3.049} \end{pmatrix}, \ \Omega_2 = \begin{pmatrix} \underset{(0.003)}{0.015} & \underset{(0.007)}{0.015} \\ \underset{(0.007)}{0.015} & \underset{(0.048)}{0.221} \end{pmatrix}$$

Wald test of Serially uncorrelated MC

W(4)=489 (0%)

Misspecification Tests

| | Equation 1 | Equation 2 |
|---|---|---|
| Autocorrelation | F(1,125)=2.7 (10%), | F(1,125)=0.01 (93%) |
| ARCH | F(4,125)=0.7 (55%), | F(4,125)=1.52 (20%) |
| Markov | F(4,125)=0.1 (98%), | F(4,125)=0.47 (76%) |

Table 3 shows that all the parameters (except $B_i^{(2,1)}$) are very close to those in table 1, but that the standard errors of the state invariant terms are an order of magnitude smaller than those in table 1 (with only one exception). The estimated state probabilities $\xi_{t|t}$ (not displayed) are also very similar. Moreover, the miss-specification tests do not suggest that this model is inappropriate, indicating that the Sola & Driffill VAR may be overparameterised. A LR test between the models 1 and 2 gives a statistic of LR(6)=2(125.0-121.9)=6.2 (40%), which cannot be rejected. In other words, we cannot reject the hypothesis that the drift and autoregressive parameters are state invariant.

What does this imply for the premium? The premium can only be constant when selected elements of the drift and autoregressive terms are state invariant, so that this supports the constant premium restriction in Sola & Driffill *under* the null of the EH. For the model in table 3, the EH implies that $B^{(2,1)} = 2$ and $B^{(2,2)} = 0$. Estimating the model in table 3 subject to these restrictions gives us a loglikelihood value of

$L_2(\tilde{\theta}) = -131.1$. A LR test of these EH restrictions on model 2 gives a statistic of LR(2)=2(131.1-125.0)=12.2 (0.2%), which is rejected at the 1% level. This suggests the possibility that the Sola & Driffill non-rejection of the EH stems from increased uncertainty due to an overparameterised VAR[3], and perhaps also the Wald test above.

**Table 4**

VAR in table 1:

$\text{Corr}\left(S_t, S_t^*\right) = 0.4 \quad \text{Var}\left[S_t\right] / \text{Var}\left[S_t^*\right] = 4.04$

VAR in table 3:

$\text{Corr}\left(S_t, S_t^*\right) = 0.65 \quad \text{Var}\left[S_t\right] / \text{Var}\left[S_t^*\right] = 3.4$

VAR based on OLS (not displayed)

$\text{Corr}\left(S_t, S_t^*\right) = 0.43 \quad \text{Var}\left[S_t\right] / \text{Var}\left[S_t^*\right] = 5.2$

Finally, we investigate whether this statistical rejection is also an "economic rejection" in the sense of Campbell & Shiller (1987). For this purpose, we compute the ex ante optimal unrestricted forecast of the spread, $S_t^* = 0.5\text{E}[\Delta R_{t+1}^{(1)} \big| \mathcal{Y}_t]$, which is given by

$$S_t^* = 0.5 e_2' \, aP\xi_{t|t} + 0.5 e_2' \, bP_{np}\tilde{Y}_t. \qquad (67)$$

To the extent that the EH is right, there should be a close correspondence between $S_t$ and $S_t^*$. They are plotted in figures 4 and 5 for the VAR:s in tables 1 and 3 respectively, which show that in state 2, the EH does a good job of capturing movements in the spread, but a bad one in state 1. This visual inspection is confirmed by the correlation and variance ratio in table 4. Moreover, the correlation between $S_t$ and $S_t^*$ only approaches 90% if we confine ourselves to the mid 1980:s and onwards. Remarkably enough, the VAR with only state dependent covariance term (in table 3) is better at predicting the spread than the more flexible VAR (in table 1). Overall the evidence of the "economic" performance of the EH is not as impressive as in Campbell & Shiller (1987) for the long end of the term structure, where the correlation was above 90 % for the whole period. Note that model 3 achieves a much better fit than a VAR based on OLS.

[3] A potential problem with this sequential testing procedure is that the true significance level may be different than those in standard statistics tables.

Figure 4. Ex Ante Optimal Prediction, Model 1

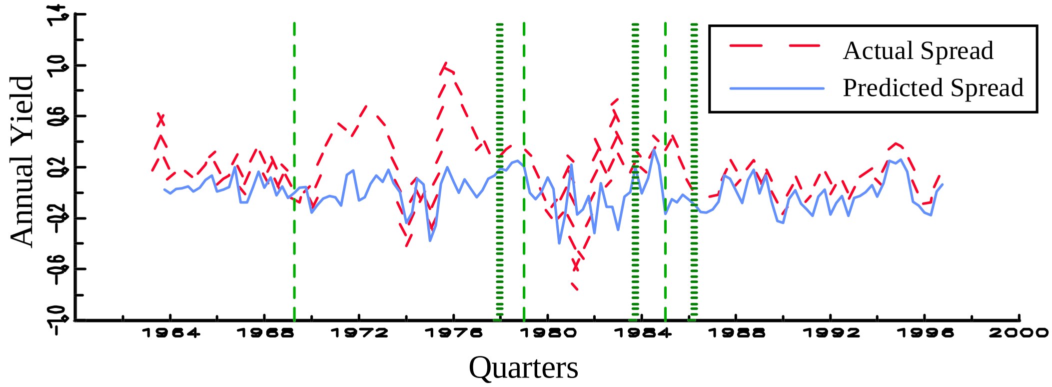
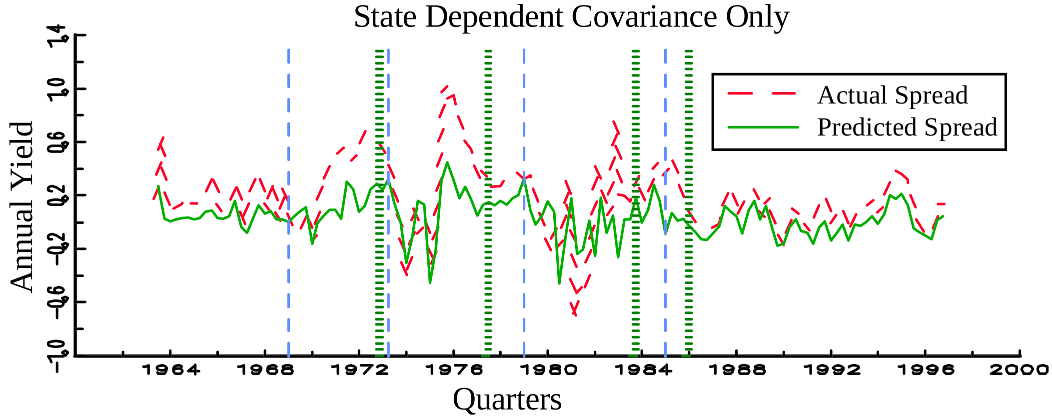



Figure 5. Ex Ante Optimal Prediction, Model 2
State Dependent Covariance Only




## 8 Concluding Remarks

VAR:s with Markov switching may sometimes be suggested as a natural statistical model from economic theory, or they may simply be better characterisation of data - a parsimonious way to capture salient features of the data. This paper shows how tests of rational expectations that have been popular in standard linear VAR:s can be generalised to allow for $q$ unobservable states. In general, the restrictions are non-linear, but in the important special case of infinite horizon models they are linear and do not depend on the transition probabilities. One other way the RE restrictions could be satisfied - discussed but not implemented in the paper - would be if a certain orthogonality condition were satisfied, with the interpretation of a serially uncorrelated Markov chain. We argued against this approach on both practical and theoretical grounds.

It was shown how the methods could be extended to allow for a state dependent discount factor with very little added complexity in the formulas. It was argued that changing the (exogenous) discount factors might be a useful diagnostic tool: determining how much a discount factor has to be changed to alter the results gives more information about robustness. Having more than one discount factor to change gives more flexibility.

Finally, the methods were applied to testing the expectations hypothesis on 3 and 6 month US bills, performing a similar test as Sola & Driffill (1994) but with a different regime shift model allowing for richer dynamics. The restrictions in this paper are different, however, in that here we obtain 4 restrictions *across* states on the autoregressive parameters, while their restrictions on the autoregressive parameters are all state invariant; the only restriction reflecting regime shifting in their paper is one that comes from the drift term.

The evidence from the tests in this paper suggests that the Sola & Driffill non-rejection of the EH may be fragile, possibly due to the use of an overparameterised VAR. We also investigated the economic fit of the model in the sense of Campbell & Shiller (1987). In state 2, the model does good job of capturing movements in the spread, but significantly worse in state 1, which covers the years when the change in operating procedures increased the volatility (and level) of interest rates (the Volcker effect). Overall, the model here provides a parsimonious representation of the data with better predictive power than a VAR based on OLS, but is not quite able to represent the years 1979-82.

## Appendix A

**Proof of Lemma 4.2**

By the Markov property,

$$\begin{aligned}\mathrm{pr}\left[s_{t+1}=i_1\middle|\mathcal{Y}_t\right]&=\sum_{i_0=1}^{q}\mathrm{pr}\left[s_{t+1}=i_1\middle|s_t=i_0,\mathcal{Y}_t\right]\mathrm{pr}\left[s_t=i_0\middle|\mathcal{Y}_t\right]\\&=\sum_{i_0=1}^{q}\mathrm{pr}\left[s_{t+1}=i_1\middle|s_t=i_0\right]\mathrm{pr}\left[s_t=i_0\middle|\mathcal{Y}_t\right].\end{aligned}\tag{A.1}$$

We can write (A.1) as

$$\begin{aligned}&\sum_{i_0=1}^{q}p_{i_0,i_1}\,\mathrm{pr}\left[s_t=i_0\middle|\mathcal{Y}_t\right]\\&=p_{1i_1}\mathrm{pr}\left[s_t=1\middle|\mathcal{Y}_t\right]+\cdots+p_{qi_1}\mathrm{pr}\left[s_t=q\middle|\mathcal{Y}_t\right]\\&=\left[p_{1i_1},\quad\cdots\quad,p_{qi_1}\right]\xi_{t|t}\\&=e_{i_1}'P\xi_{t|t}.\end{aligned}\tag{A.2}$$

Similarly,

$$\begin{aligned}\mathrm{pr}\left[s_{t+2}=i_2,s_{t+1}=i_1\middle|\mathcal{Y}_t\right]&=p_{i_1i_2}\left(e_{i_1}'P\,\xi_{t|t}\right)\\&=\left(e_{i_2}'P\,e_{i_1}\right)\left(e_{i_1}'P\,\xi_{t|t}\right),\end{aligned}\tag{A.3}$$

and so on. □

**Proof of Lemma 4.3**

First, substitute (19) and (20) into (21), and for $j\geq 2$ we get

$$\begin{aligned}\mathrm{E}\left[y_{t+j}\middle|\mathcal{Y}_t\right]=\sum_{i_j=1}^{q}\cdots\sum_{i_1=1}^{q}\left\{\mu_{i_j}+J\sum_{h=1}^{j-1}\left(\prod_{m=1}^{h}B_{i_{j+1-m}}\right)J'\,\mu_{i_{j-h}}+J\left(\prod_{m=1}^{j}B_{i_{j+1-m}}\right)Y_t\right\}\times\\\left(\prod_{f=0}^{j-2}e_{i_{j-f}}'\,Pe_{i_{j-f-1}}\right)\left(e_{i_1}'\,P\,\xi_{t|t}\right).\end{aligned}\tag{A.4}$$

**Step 1.** First we consider the term in (A.4) that involves $Y_t$. When $j=3$ this term is

$$J\sum_{i_3=1}^{q}\sum_{i_2=1}^{q}\sum_{i_1=1}^{q}B_{i_3}B_{i_2}B_{i_1}Y_t\left(e_{i_3}'\,P\,e_{i_2}\right)\left(e_{i_2}'\,P\,e_{i_1}\right)\left(e_{i_1}'\,P\,\xi_{t|t}\right),\tag{A.5}$$

which can be written as

$$J\sum_{i_3=1}^{q}B_{i_3}\sum_{i_2=1}^{q}\left(e_{i_3}'\,P\,e_{i_2}\right)B_{i_2}\left[\sum_{i_1=1}^{q}\left(e_{i_2}'\,P\,e_{i_1}\right)B_{i_1}\left(e_{i_1}'\,P\,\xi_{t|t}\right)Y_t\right].\tag{A.6}$$

Recall that $\tilde{Y}_t=\left(\xi_{t|t}\otimes Y_t\right)$, and note that the term inside the square brackets in (A.6) is

$$\left[\sum_{i_1=1}^{q}\left(e_{i_2}' P e_{i_1}\right)B_{i_1}\left(e_{i_1}' P \xi_{t|t}\right)Y_t\right]=\left[\left(e_{i_2}' P e_1\right)B_1\left(e_1' P\xi_{t|t}\right)+\cdots+\left(e_{i_2}' P e_q\right)B_q\left(e_q' P\xi_{t|t}\right)\right]Y_t$$

$$=\left[\left(e_{i_2}' P e_1\right)B_1,\cdots,\left(e_{i_2}' P e_q\right)B_q\right]\begin{bmatrix}\left(e_1' P\xi_{t|t}\right)I_{np}\\ \vdots \\ \left(e_q' P\xi_{t|t}\right)I_{np}\end{bmatrix}Y_t$$

$$=\left[\left(e_{i_2}' P e_1\right)I_{np},\cdots,\left(e_{i_2}' P e_q\right)I_{np}\right]B\left[\begin{pmatrix}e_1'\\ \vdots \\ e_q'\end{pmatrix}P\xi_{t|t}\otimes I_{np}\right]Y_t$$

$$=\left[\left(e_{i_2}' P e_1,\cdots,e_{i_2}' P e_q\right)\otimes I_{np}\right]B\left(P\otimes I_{np}\right)\tilde{Y}_t \qquad \text{(A.7)}$$

$$=\left[\left(e_{i_2}' P\left[e_1,\cdots,e_q\right]\right)\otimes I_{np}\right]B\,P_{np}\tilde{Y}_t$$

$$=\left(e_{i_2}'\otimes I_{np}\right)\left(P\otimes I_{np}\right)BP_{np}\tilde{Y}_t$$

$$=\left(e_{i_2}'\otimes I_{np}\right)\Phi\, P_{np}\tilde{Y}_t.$$

Substitute this expression back into (A.6), and we obtain

$$J\sum_{i_3=1}^{q} B_{i_3}\left[\sum_{i_2=1}^{q}\left(e_{i_3}' P e_{i_2}\right)B_{i_2}\left(e_{i_2}'\otimes I_{np}\right)\right]\Phi\, P_{np}\tilde{Y}_t. \qquad \text{(A.8)}$$

Again consider the expression inside the square brackets:

$$\sum_{i_2=1}^{q}\left(e_{i_3}' P e_{i_2}\right)B_{i_2}\left(e_{i_2}'\otimes I_{np}\right)=$$

$$=\left[\left(e_{i_3}' P e_1\right)B_1\left(e_1'\otimes I_{np}\right)+\cdots+\left(e_{i_3}' P e_q\right)B_q\left(e_q'\otimes I_{np}\right)\right]$$

$$=\left[\left(e_{i_3}' P e_1\right)B_1,\cdots,\left(e_{i_3}' P e_q\right)B_q\right]\begin{bmatrix}e_1'\otimes I_{np}\\ \vdots \\ e_q'\otimes I_{np}\end{bmatrix} \qquad \text{(A.9)}$$

$$=\left(e_{i_3}'\otimes I_{np}\right)\Phi.$$

Substitute this back into (A.8) and we obtain

$$J\sum_{i_3=1}^{q} B_{i_3}\left(e_{i_3}'\otimes I_{np}\right)\Phi^2 P_{np}\tilde{Y}_t = J\left[B_1\left(e_1'\otimes I_{np}\right)+\cdots+B_q\left(e_q'\otimes I_{np}\right)\right]\Phi^2 P_{np}\tilde{Y}_t$$

$$= J\left[B_1,\cdots,B_q\right]\Phi^2 P_{np}\tilde{Y}_t \qquad \text{(A.10)}$$

$$= b\,\Phi^2 P_{np}\tilde{Y}_t.$$

From this we can infer that for $j\geq 2$ this term will have the form

$$b\,\Phi^{j-1}P_{np}\tilde{Y}_t. \qquad \text{(A.11)}$$

A similar calculation for $j=1$ yields

$$b\,P_{np}\tilde{Y}_t. \qquad \text{(A.12)}$$

**Step 2.** Consider $j=4$. The terms in (A.4) that do not involve $Y_t$ explicitly are

$$\sum_{i_4=1}^{q}\sum_{i_3=1}^{q}\sum_{i_2=1}^{q}\sum_{i_1=1}^{q}\left\{\mu_{i_4}+JB_{i_4}J'\mu_{i_3}+JB_{i_4}B_{i_3}J'\mu_{i_2}+JB_{i_4}B_{i_3}B_{i_2}J'\mu_{i_1}\right\}\times \left(e_{i_4}'Pe_{i_3}\right)\left(e_{i_3}'Pe_{i_2}\right)\left(e_{i_2}'Pe_{i_1}\right)\left(e_{i_1}'P\xi_{t|t}\right). \tag{A.13}$$

Let us focus on the first term in (A.13), which after some rearrangement is given by

$$\sum_{i_4=1}^{q}\mu_{i_4}\sum_{i_3=1}^{q}\left(e_{i_4}'Pe_{i_3}\right)\sum_{i_2=1}^{q}\left(e_{i_3}'Pe_{i_2}\right)\left[\sum_{i_1=1}^{q}\left(e_{i_2}'Pe_{i_1}\right)\left(e_{i_1}'P\xi_{t|t}\right)\right]. \tag{A.14}$$

Expanding the expression inside the square brackets we obtain

$$\begin{aligned}\sum_{i_1=1}^{q}\left(e_{i_2}'Pe_{i_1}\right)\left(e_{i_1}'P\xi_{t|t}\right)&=\left[e_{i_2}'Pe_1e_1'+\cdots+e_{i_2}'Pe_qe_q'\right]P\xi_{t|t}\\&=e_{i_2}'P^2\xi_{t|t}.\end{aligned} \tag{A.15}$$

Substitute (A.15) back into (A.14), and we get

$$\sum_{i_4=1}^{q}\mu_{i_4}\sum_{i_3=1}^{q}\left(e_{i_4}'Pe_{i_3}\right)\sum_{i_2=1}^{q}\left(e_{i_3}'Pe_{i_2}\right)e_{i_2}'P^2\xi_{t|t}. \tag{A.16}$$

The summations across $i_2$ and $i_3$ are done similarly, and we proceed directly to the last summation to obtain

$$\begin{aligned}\sum_{i_4=1}^{q}\mu_{i_4}e'_{i_4}P^4\xi_{t|t}&=\left[\mu_1e_1'+\cdots+\mu_qe_q'\right]P^4\xi_{t|t}\\&=aP^4\xi_{t|t}.\end{aligned} \tag{A.17}$$

From this we readily infer that for $j\geq 1$ terms of this type is given by

$$aP^j\xi_{t|t}. \tag{A.18}$$

The other terms in (A.13) are done similarly (proof available upon request), which gives

$$aP^4\xi_{t|t}+\left(b\Psi P^3\xi_{t|t}+b\Phi\Psi P^2\xi_{t|t}+b\Phi^2\Psi P\xi_{t|t}\right). \tag{A.19}$$

We see that (A.19) can be written as

$$aP^4\xi_{t|t}+b\sum_{m=0}^{2}\Phi^m\Psi P^{3-m}\xi_{t|t}. \tag{A.20}$$

Now, from (A.20) we can infer that the pattern for $j\geq 2$ is

$$aP^j\xi_{t|t}+b\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\xi_{t|t}. \tag{A.21}$$

The calculation for $j=1$ is straightforward, and omitted. □

**Proof of Proposition 4.1.**

First note that $N_0y_t=N_0JY_t=N_0JC_{np}\tilde{Y}_t$, and using the result in (A.11), (A.21), and (18) we see that the rational expectations restrictions in (9) can be written compactly as

$$\Xi^{(k)}P_{np}\left(e_{s_t}\otimes Y_t\right)+\Lambda^{(k)}e_{s_t}=\lambda'e_{s_t}, \tag{A.22}$$

where we have used the notation in proposition 4.1. We can write this system of equations as

$$\left[\Xi^{(k)}P_{np} \quad \Lambda^{(k)} - \lambda'\right]\begin{bmatrix}\left(e_{s_t} \otimes Y_t\right) \\ e_{s_t}\end{bmatrix} = 0. \tag{A.23}$$

The only expression that is orthogonal to $\left[\left(e_{s_t} \otimes Y_t\right)' \quad e_{s_t}'\right]'$ for all $t$ is the zero matrix, which shows that under the null hypothesis in (9) the restrictions in the proposition must hold . By the law of iterated expectations, these restrictions should also hold for (10). □

## Appendix B

We now state two useful propositions, and introduce a new lemma, which will be of use below.

**Proposition B.1**

Let $A$ be $m\times n$ and $B$ be $p\times q$. Then

$$\text{vec}(A\otimes B) = \left(I_n \otimes k_{q,m} \otimes I_p\right)(\text{vec}\, A\otimes \text{vec}\, B).$$

Proof: See Magnus and Neudecker (1988, p 47).

**Proposition B.2**

Let $X$ be a $n\times n$ matrix and $F(X) = X^{\tau}$. Then $\mathrm{d}F = \sum_{m=1}^{\tau} X^{m-1}\mathrm{d}X\, X^{\tau-m}$.

Proof: See Magnus and Neudecker (1988, p 183).

**Lemma B.1** (generalisation of theorem 7.1 in Magnus (1989, p 109))

Define the $rsq^2\times rsq$ matrix $\Delta_{r,s} \equiv \left(I_q \otimes k_{s,q} \otimes I_r\right)\sum_{i=1}^{q}\left((\text{vec}\, E_{ii})e_i'\otimes I_{rs}\right)$, where $E_{ij} = e_i\, e_j'$. When $r = s,$ we simply write $\Delta_r$. Let the $rq\times sq$ matrix

$$Z = \begin{pmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_q \end{pmatrix}, \tag{B.1}$$

where the diagonal matrices $Z_\tau$ are $r\times s$, and let $w(Z) = \left[Z_1 \cdots Z_q\right]$. Then we have that

$$\text{vec}\, Z = \Delta_{r,s}\text{vec}\, w(Z). \tag{B.2}$$

**Proof:** We can write $Z = \sum_{i=1}^{q}\left(E_{ii} \otimes Z_i\right)$. Apply the vec operator on both sides and we obtain

$$\begin{aligned}\operatorname{vec}(Z) &= \textstyle\sum_{i=1}^{q} \operatorname{vec}\left(E_{ii} \otimes Z_i\right) \\ &= \left(I_q \otimes k_{s,q} \otimes I_r\right) \textstyle\sum_{i=1}^{q} \left(\operatorname{vec} E_{ii} \otimes \operatorname{vec} Z_i\right) \qquad \text{(B.3)} \\ &= \left(I_q \otimes k_{s,q} \otimes I_r\right) \textstyle\sum_{i=1}^{q} \left(\operatorname{vec} E_{ii} \otimes I_{rs}\right) \operatorname{vec} Z_i .\end{aligned}$$

Now, we can write $Z_i = w(Z)(e_i \otimes I_s)$. Using this expression in (B.3) we obtain

$$\begin{aligned}\operatorname{vec}(Z) &= \left(I_q \otimes k_{s,q} \otimes I_r\right) \textstyle\sum_{i=1}^{q} \left(\operatorname{vec}(E_{ii}) \otimes I_{rs}\right) \operatorname{vec}\left(w(Z)(e_i \otimes I_s)\right) \qquad \text{(B.4)} \\ &= \left(I_q \otimes k_{s,q} \otimes I_r\right) \left[\textstyle\sum_{i=1}^{q} \left(\operatorname{vec}(E_{ii}) e'_i \otimes I_{rs}\right)\right] \operatorname{vec}(w(Z)).\end{aligned}$$

□

**Proof of Proposition 6.1**

From proposition 4.1, for $j \geq 2$ we have

$$\Xi_j = N_j \, b \, \Phi^{j-1}. \qquad \text{(B.5)}$$

The differential is

$$\begin{aligned}\operatorname{dvec} \Xi_j &= \operatorname{vec}\left[N_j \mathrm{d}b \, \Phi^{j-1}\right] \qquad \text{(B.6)} \\ &\quad + \operatorname{vec}\left[N_j b \, \mathrm{d}[\Phi^{j-1}]\right].\end{aligned}$$

For the first term on the RHS of (B.6) we immediately obtain

$$\left((\Phi^{j-1})' \otimes N_j\right) \mathrm{d}\beta. \qquad \text{(B.7)}$$

The second term on the RHS of (B.6) is

$$\left(I_{npq} \otimes N_j b\right) \operatorname{vec} \mathrm{d}[\Phi^{j-1}] = \textstyle\sum_{m=1}^{j-1} \left((\Phi^{j-1-m})' \otimes N_j b \Phi^{m-1}\right) \operatorname{dvec}\Phi, \qquad \text{(B.8)}$$

where we have used proposition B.2. Further,

$$\begin{aligned}\operatorname{dvec} \Phi &= \operatorname{vec}\left[\left(\mathrm{d}P \otimes I_{np}\right) B\right] + \operatorname{vec}\left[\left(P \otimes I_{np}\right) \mathrm{d}B\right] \qquad \text{(B.9)} \\ &= \left(B' \otimes I_{npq}\right)\left(I_q \otimes k_{np,q} \otimes I_{np}\right)\left(I_{q^2} \otimes \operatorname{vec} I_{np}\right)\left(\operatorname{vec} \mathrm{d}P \otimes 1\right) + \left(I_{npq} \otimes P_{np}\right) \operatorname{dvec} B,\end{aligned}$$

where we have used proposition B.1. Now, applying lemma B.1, this can be written as

$$\operatorname{dvec} \Phi = \left(B' \otimes I_{npq}\right) \kappa_{np} \Upsilon \, \mathrm{d}\gamma + \left(I_{npq} \otimes P_{np}\right) \Delta_{np} \operatorname{vec} w(\mathrm{d}B), \qquad \text{(B.10)}$$

where

$$\begin{aligned}\operatorname{vec} w(\mathrm{d}B) &= \operatorname{vec}\left(J' \, \mathrm{d}b\right) \qquad \text{(B.11)} \\ &= \left(I_{npq} \otimes J'\right) \mathrm{d}\beta.\end{aligned}$$

Substitute (B.11) into (B.10), and we get

$$\operatorname{dvec} \Phi = \left(B' \otimes I_{npq}\right) \kappa_{np} \Upsilon \, \mathrm{d}\gamma + \left(I_{npq} \otimes P_{np}\right) \Delta_{np} \left(I_{npq} \otimes J'\right) \mathrm{d}\beta. \qquad \text{(B.12)}$$

Substitute this expression into (B.8), and we get

$$\sum_{m=1}^{j-1}\left((B\Phi^{j-1-m})'\otimes N_j b\Phi^{m-1}\right)\kappa_{np}\Upsilon\,\mathrm{d}\gamma$$
$$+\sum_{m=1}^{j-1}\left((\Phi^{j-1-m})'\otimes N_j b\Phi^{m-1}P_{np}\right)\Delta_{np}\left(I_{npq}\otimes J'\right)\mathrm{d}\beta. \quad \text{(B.13)}$$

Expression (B.13) gives us the differential of the second term on the RHS of (B.6). Collecting the terms in (B.7), and (B.13) we get for $j\geq 2$,

$$\mathrm{dvec}\,\Xi_j=\left((\Phi^{j-1})'\otimes N_j\right)\mathrm{d}\beta$$
$$+\sum_{m=1}^{j-1}\left((B\Phi^{j-1-m})'\otimes N_j b\Phi^{m-1}\right)\kappa_{np}\Upsilon\,\mathrm{d}\gamma \quad \text{(B.14)}$$
$$+\sum_{m=1}^{j-1}\left((\Phi^{j-1-m})'\otimes N_j b\Phi^{m-1}P_{np}\right)\Delta_{np}\left(I_{npq}\otimes J'\right)\mathrm{d}\beta,$$

from which the required derivatives are easily obtained. For $j=1$ we have

$$\mathrm{dvec}\,\Xi_1=\mathrm{vec}\left[N_1\mathrm{d}b\,P_{np}\right]$$
$$=\left(P'_{np}\otimes N_1\right)\mathrm{d}\beta. \quad \text{(B.15)}$$

□

**Proof of Proposition 6.2**

We know turn to the differential of $\mathrm{vec}\,\Lambda_j$, and obtain for $j\geq 3$

$$\mathrm{dvec}\,\Lambda_j=\mathrm{dvec}\left[N_j a\,P^j\right]+\mathrm{dvec}\left[N_j b\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right]. \quad \text{(B.16)}$$

Expanding the first term on the RHS of (B.16) we get

$$\mathrm{dvec}\left[N_j a\,P^j\right]=\mathrm{vec}\left[N_j\mathrm{d}a\,P^j\right]+\mathrm{vec}\left[N_j a\,\mathrm{d}[P^j]\right]$$
$$=\left((P^j)'\otimes N_j\right)\mathrm{d}\alpha+\sum_{m=1}^{j}\left((P^{j-m})'\otimes N_j a\,P^{m-1}\right)\Upsilon\,\mathrm{d}\gamma. \quad \text{(B.17)}$$

Next, we expand the second term on the RHS of (B.16).

$$\mathrm{dvec}\left[N_j b\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right]=\mathrm{vec}\left[N_j\mathrm{d}b\sum_{m=0}^{j-2}\Phi^m\Psi P^{j-1-m}\right]$$
$$+\mathrm{vec}\left[N_j b\sum_{m=1}^{j-2}\mathrm{d}[\Phi^m]\Psi P^{j-1-m}\right]$$
$$+\mathrm{vec}\left[N_j b\sum_{m=0}^{j-2}\Phi^m\mathrm{d}\Psi\,P^{j-1-m}\right] \quad \text{(B.18)}$$
$$+\mathrm{vec}\left[N_j b\sum_{m=0}^{j-2}\Phi^m\Psi\,\mathrm{d}[P^{j-1-m}]\right].$$

The first term on the RHS of (B.18) is

$$\sum_{m=0}^{j-2}\left[\left(\Phi^m\Psi P^{j-1-m}\right)'\otimes N_j\right]\mathrm{d}\beta, \quad \text{(B.19)}$$

while the second term is

$$\sum_{m=1}^{j-2}\left((\Psi P^{j-1-m})'\otimes N_j b\right)\mathrm{vec}\left[\sum_{s=1}^{m}\Phi^{s-1}\mathrm{d}\Phi\;\Phi^{m-s}\right]=$$
$$\sum_{m=1}^{j-2}\sum_{s=1}^{m}\left((\Phi^{m-s}\Psi P^{j-1-m})'\otimes N_j b\Phi^{s-1}\right)\mathrm{dvec}\,\Phi, \quad \text{(B.20)}$$

and we already know $\mathrm{dvec}\Phi$ from (B.12). The third term in (B.18) is

$$\operatorname{vec}\left[N_j b \sum_{m=0}^{j-2} \Phi^m \mathrm{d}\Psi\, P^{j-1-m}\right] = \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right) \mathrm{dvec}\Psi. \qquad \text{(B.21)}$$

Now,

$$\mathrm{dvec}\Psi = \left(\mu' \otimes I_q \otimes J'\right) \kappa_n \Upsilon\, \mathrm{d}\gamma + \left(I_q \otimes P \otimes J'\right) \mathrm{dvec}\mu \qquad \text{(B.22)}$$

and

$$\mathrm{dvec}\,\mu = \Delta_{n,1} \operatorname{vec} w(\mathrm{d}\mu) = \Delta_{n,1}\, \mathrm{d}\alpha \qquad \text{(B.23)}$$

So that (B.21) becomes

$$\begin{aligned} &\sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right)\left(\mu' \otimes I_q \otimes J'\right) \kappa_n \Upsilon\, \mathrm{d}\gamma \\ &+ \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right)\left(I_q \otimes P \otimes J'\right) \Delta_{n,1}\, \mathrm{d}\alpha. \end{aligned} \qquad \text{(B.24)}$$

Finally, the fourth term in (B.18) is

$$\begin{aligned} \operatorname{vec}\left[N_j b \sum_{m=0}^{j-2} \Phi^m \Psi\, \mathrm{d}[P^{j-1-m}]\right] &= \sum_{m=0}^{j-2} \left(I_q \otimes N_j b \Phi^m \Psi\right) \operatorname{vec}\left[\sum_{s=1}^{j-1-m} P^{s-1} \mathrm{d}P\, P^{j-1-m-s}\right] \\ &= \sum_{m=0}^{j-2} \sum_{s=1}^{j-1-m} \left((P^{j-1-m-s})' \otimes N_j b \Phi^m \Psi P^{s-1}\right) \Upsilon\, \mathrm{d}\gamma. \end{aligned} \qquad \text{(B.25)}$$

Collecting the terms in (B.17), (B.19), (B.20), (B.24), and (B.25), we get for $j \geq 3$

$$\begin{aligned} \mathrm{dvec}\,\Lambda_j = &\left((P^j)' \otimes N_j\right) \mathrm{d}\alpha + \sum_{m=1}^{j} \left((P^{j-m})' \otimes N_j a P^{m-1}\right) \Upsilon\, \mathrm{d}\gamma \\ &+ \sum_{m=0}^{j-2} \left[\left(\Phi^m \Psi P^{j-1-m}\right)' \otimes N_j\right] \mathrm{d}\beta \\ &+ \sum_{m=1}^{j-2} \sum_{s=1}^{m} \left((B \Phi^{m-s} \Psi P^{j-1-m})' \otimes N_j b \Phi^{s-1}\right) \kappa_{np} \Upsilon\, \mathrm{d}\gamma \\ &+ \sum_{m=1}^{j-2} \sum_{s=1}^{m} \left((\Phi^{m-s} \Psi P^{j-1-m})' \otimes N_j b \Phi^{s-1} P_{np}\right) \Delta_{np} \left(I_{npq} \otimes J'\right) \mathrm{d}\beta \\ &+ \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right)\left(\mu' \otimes I_q \otimes J'\right) \kappa_n \Upsilon\, \mathrm{d}\gamma \\ &+ \sum_{m=0}^{j-2} \left((P^{j-1-m})' \otimes N_j b \Phi^m\right)\left(I_q \otimes P \otimes J'\right) \Delta_{n,1}\, \mathrm{d}\alpha \\ &+ \sum_{m=0}^{j-2} \sum_{s=1}^{j-1-m} \left((P^{j-1-m-s})' \otimes N_j b \Phi^m \Psi P^{s-1}\right) \Upsilon\, \mathrm{d}\gamma, \end{aligned} \qquad \text{(B.26)}$$

which easily yields the required derivatives. For $j = 2$, there is no $\Phi$ term in (B.16) and so the terms from (B.20) do not enter into $\mathrm{dvec}\Lambda_2$, which is given by

$$\begin{aligned} \mathrm{dvec}\Lambda_2 = &\left((P^2)' \otimes N_2\right) \mathrm{d}\alpha + \sum_{m=1}^{2} \left((P^{2-m})' \otimes N_2 a P^{m-1}\right) \Upsilon\, \mathrm{d}\gamma \\ &+ \left((\Psi P)' \otimes N_2\right) \mathrm{d}\beta + \left(P' \otimes N_2 b\right)\left(\mu' \otimes I_q \otimes J'\right) \kappa_n \Upsilon\, \mathrm{d}\gamma \\ &+ \left(P' \otimes N_2 b\right)\left(I_q \otimes P \otimes J'\right) \Delta_{n,1} \mathrm{d}\alpha + \left(I_q \otimes N_2 b \Psi\right) \Upsilon\, \mathrm{d}\gamma. \end{aligned} \qquad \text{(B.27)}$$

For $j = 1$, we get the simpler expression

$$\begin{aligned} \mathrm{dvec}\Lambda_1 &= \operatorname{vec}\left[N_1 \mathrm{d}a\, P\right] + \operatorname{vec}\left[N_1 a\, \mathrm{d}P\right] \\ &= \left(P' \otimes N_1\right) \mathrm{d}\alpha + \left(I_q \otimes N_1 a\right) \Upsilon\, \mathrm{d}\gamma. \end{aligned} \qquad \text{(B.28)}$$

□

## References

Baillie, Richard T. (1989), “Econometric Tests of Rationality and Market Efficiency”, *Econometric Reviews*, 8(2), pp 151-86.

Bickel, Peter J., Ya’acov Ritov, and Tobias Rydén (1997), “Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models”, *mimeo*, Dept. of Mathematical Statistics, Lund University, Sweden.

Blix, Mårten (1996), “Term Premia Under Switching Regimes”, *Sveriges Riksbank* working paper No 27.

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

Engel, Charles and James Hamilton (1990), “Long Swings in the Dollar: Are They in the Data and Do Markets Know It?”, *American Economic Review*, Vol. 80, No. 4, pp 689-713.

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.

Hamilton, James D. and Raul Susmel (1994), “Autoregressive Conditional Heteroskedasticity and Changes in Regime”, *Journal of Econometrics*, 64, pp 307-333.

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.

Karlsen, H. (1990), *A Class of Non-linear Time Series Models,* PhD dissertation, Department of Mathematics, University of Bergen, Norway.

Krishnamurthy, Vikram and Tobias Rydén (1996), "Consistent Estimation of Linear and Non-linear Autoregressive Models with Markov Regime", *mimeo*, Dept. of Mathematical Statistics, Lund University, Sweden.

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

Magnus, Jan R. (1988), *Linear Structures* (Griffin's Statistical Monographs and Courses 42, Charles Griffin & Company: London).

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.

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.

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 (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.