SWITCHING REGIME ESTIMATION.(**CONTINUATION**)

Specification Tests.

Hamilton (1990) proposes many alternative specification tests to evaluate the Markov Switching Model. In this notes we will only focus on specification tests based on the properties of the standardized residuals.

To construct these residuals we first proceed to calculate the conditional expectation of y_t given information at time $t-1$.

$$\begin{aligned} E(y_t|I_{t-1}) &= \alpha_0 + \alpha_1 E(x_t|I_{t-1}) + \phi_1(y_{t-1} - \alpha_1 E(x_{t-1}|I_{t-1}) - \alpha_0) \\ &\quad + ... + \phi_m(y_{t-m} - \alpha_1 E(x_{t-m}|I_{t-1}) - \alpha_0), \end{aligned}$$

where $E(x_{t-m}|I_{t-1}) = P(x_{t-m} = 1|I_{t-1})$. for $m > 0$. and

$$E(x_t|I_{t-1}) = \frac{(1-q)}{(2-p-q)} + (P(x_{t-1}|I_{t-1}) - \frac{(1-q)}{(2-p-q)})(p+q-1) \ for \ m = 0.$$

(Notice that we assume that the public does not observe the state and therefore we substitute x_{t-1} by $P(x_{t-1}|I_{t-1})$).

Also note that the probabilities $P(x_{t-m}|I_{t-1})$, for $m > 1$, are called "smoothing probabilities" and can easily be calculated from the "filtering probabilities" i.e. when $0 \leq m \leq 1$. Then the residuals are just $\varepsilon_t = y_t - E(y_t|I_{t-1})$.

To compute the standardized residuals we need to calculate the conditional standard deviation.

We proceed in (3) steps

(1) First make use of the autoregressive representation of the Markov process

$$x_t = (1-q) + (-1+p+q)x_{t-1} + \zeta_{2,t}$$

For this process the error, conditional on $x_{t-1} = 1$, can be characterized as

$$\zeta_{2,t} = \begin{array}{l} (1-p) \text{ with probability } p \\ -p \text{ with probability } 1-p \end{array}$$

and conditional on $x_{t-1} = 0$.

$$\zeta_{2,t} = \begin{array}{l} -(1-q) \text{ with probability } q \\ q \text{ with probability } 1-q \end{array}$$

2) Calculate the variance of the error term, $\zeta_{2,t}$, conditional on the state at $t-1$.

$$\begin{aligned} E(\zeta_{2,t}^2|x_{t-1} &= 1) = (1-p)^2 p + p^2(1-p) = p(1-p) \\ E(\zeta_{2,t}^2|x_{t-1} &= 0) = (1-q)^2 q + q^2(1-q) = q(1-q) \end{aligned}$$

3) Calculate the conditional variance (conditional on $I_{t-1} = \{y_{t-1}, ...y_0\}$)

We start by calculating the state dependent variance $\sigma_{x_t}^2$ as a function of the Markov switching parameters.

Conditional on $x_{t-1} = 1$, the switching variance can be written as:

$$\sigma_{x_t}^2 = E(\sigma_{x_t}^2|x_{t-1} = 1) + V(\mu_{x_t}|x_{t-1} = 1)$$

where

a)

$$E(\sigma_{x_t}^2|x_{t-1} = 1) = (E(\sigma_{x_t}||x_{t-1} = 1))^2 + Var((\sigma_{x_t}|x_{t-1} = 1))$$

(since σ_{x_t} is a random variable) and

$$Var((\sigma_{x_t}|x_{t-1} = 1)) = E(\sigma_{x_t}^2|x_{t-1} = 1) - (E(\sigma_{x_t}|x_{t-1} = 1))^2$$

Then using that

(iia)

$$(E(\sigma_{x_t}|x_{t-1} = 1))^2 = (w_0 + w_1 E(x_t|x_{t-1} = 1)^2 = (w_0 + w_1 p)^2$$

$$Var((\sigma_{x_t}|x_{t-1} = 1)) = Var(w_0 + w_1 x_t|x_{t-1} = 1) = w_1^2 p(1-p).$$

$$\begin{aligned} V(\mu_{x_t}|x_{t-1} &= 1) = E\{(\mu_{x_t} - E(\mu_{x_t}))^2|x_{t-1} = 1\} \\ &= E\{(\alpha_0 + \alpha_1 x_t - E(\alpha_0 + \alpha_1 x_t))^2|x_{t-1} = 1\} \\ &= \alpha_1^2\{E(x_t - E(x_t))^2|x_{t-1} = 1\} = \alpha_1^2 p(1-p) \end{aligned}$$

Collecting all these terms we can see that

$$\sigma_{x_t}^2 = (w_0 + w_1 p)^2 + w_1^2 p(1-p) + \alpha_1^2 p(1-p)$$

We can obtain a similar formulae for the variance conditional on $x_{t-1} = 0$.

$$\sigma_{x_t}^2 = E(\sigma_{x_t}^2|x_{t-1} = 0) + V(\mu_{x_t}|x_{t-1} = 0)$$

and doing the same transformations for state 0 we obtain

$$\sigma_{x_t}^2 = (w_0 + w_1(1-q))^2 + w_1^2 q(1-q) + \alpha_1^2 q(1-q).$$

Clearly the state is not observed at time $t-1$ but we can use the filtered probabilities to make an inference of the unobserved state. Then the conditional variance (on information on time $t-1$) is

$$\begin{aligned}\sigma_t^2 &= ((w_0 + w_1 p)^2 + w_1^2 p(1-p) + \alpha_1^2 p(1-p))P(x_{t-1} = 1|I_{t-1}) \\ &+ ((w_0 + w_1(1-q))^2 + w_1^2 q(1-q) + \alpha_1^2 q(1-q))(1 - P(x_{t-1} = 1|I_{t-1})).\end{aligned}$$

Then the standardized residuals are simply, $v_t = \varepsilon_t/\sigma_t$ and we may conduct standard specification tests for these residuals.

Number of States and Specification tests.

Crucial to the Hamilton methodology is to rightly identify the number of states or regimes. This can not be done by standard econometrics techniques, i.e. testing whether α_1 and w_1 are both equal to zero, since under this assumption, p and q are not identified and the distribution is not standard. Hamilton proposes to use simple specification tests as a mean of assessing whether the estimated equation contains the right number of states. If the data has, say 3 "primitive" states, and we estimated a 2 states Markov process, then the estimated model should have misspecified residuals.

Which Parameters are allowed to Switch.

In principle the Hamilton filter is general enough to allow all the parameters to switch. Nevertheless by doing this, is easy to end with non-identified models. A possible strategy can be to choose alternative switching parameterizations and see which one is favored by the data.

Rational Expectations and Regime Changes.

Changes in regime have attracted the attention of economists in a number of fields. The widespread adoption of the rational expectations hypothesis has made this issue even more important. As is well known, according to this hypothesis, expectations are forward-looking. Expected future changes in a policy variable (e.g. the money supply) will then affect the current value of the variable from which expectations have been formed (e.g. the price level or exchange rate in a monetary model). This issue has been examined in various contexts. For example, in the literature on speculative bubbles Blanchard (1979), Flood and Garber (1980) and Hamilton (1986) have shown that a bubble cannot be distinguished from a fundamental solution with an expected change in regime. Similarly in the literature on self-fulfilling speculative attacks, where if a sufficiently large group of speculators believe that the government will change the policy from, say, zero domestic credit expansion to inflationary finance, a collapse in the exchange rate regime could then occur solely because the public believes that the regime will change. This list of examples could clearly be extended. Note also that the issue is important for the definition of rational expectations: if agents do not take into account changes in regime, their rational expectations forecasts would be consistently under or over-estimating the projected variable.

A BIVARIATE VAR MODEL WITH REGIME-SWITCHING.

We consider a VAR process in two variables, with m lags, with the feature that the means of each equation and the variance-covariance matrix are allowed to switch endogenously between two possible states. The two equations that define the VAR are influenced by the same state variable. The state is not observed and has to be inferred from a filter. Therefore, we consider the following centered bivariate autoregressive process:

$$S'_t = \Phi_1 S'_{t-1} + ... + \Phi_m S'_{t-m} + \psi_1 D'_{t-1} + ... + \psi_m D'_{t-m} + (\omega_0 + \omega_1 x_t)\nu_t$$

$$D'_t = \varphi_1 S'_{t-1} + ... + \varphi_m S'_{t-m} + \Omega_1 D'_{t-1} + ... + \Omega_m D'_{t-m} + (\tau_0 + \tau_1 x_t)\varepsilon_t$$

where the centered variables are defined by the two following equations:

$$S'_t = S_t - \alpha_0 - \alpha_1 x_t$$

$$D'_t = D_t - \beta_0 - \beta_1 x_t$$

A prime (') is used to denote centred variables in the remainder of the paper. x_t denotes the unobserved state of the system and takes values 0 and 1. x_t is governed by a Markov process, summarized by the probabilities: $prob(x_t = 0|x_{t-1} = 0) = q$ and $prob(x_t = 1|x_{t-1} = 1) = p$. Substituting the centered variables into the VAR and rearranging terms, we obtain the following expression for S_t and D_t .

$$\begin{aligned} S_t \quad = \quad & \alpha_0(1 - \Phi_1 - . - \Phi_m) + \beta_0(-\psi_1 - . - \psi_m) \\ & +\Phi_1 S_{t-1} + . + \Phi_m S_{t-m} + \psi_1 D_{t-1} + . + \psi_m D_{t-m} \\ & +\alpha_1(x_t - \Phi_1 x_{t-1} - . - \Phi_m x_{t-m}) \\ & +\beta_1(-\psi_1 x_{t-1} - . - \psi_m x_{t-m}) + (\omega_0 + \omega_1 x_t)\nu_t \end{aligned}$$

$$
\begin{aligned}
D_t &= \alpha_0(-\varphi_1 - . - \varphi_m) + \beta_0(1 - \Omega_1 - . - \Omega_m) \\
&\quad +\varphi_1 S_{t-1} + . + \varphi_m S_{t-m} + \Omega_1 D_{t-1} + . + \Omega_m D_{t-m} \\
&\quad +\alpha_1(-\varphi_1 x_{t-1} - . - \varphi_m x_{t-m}) \\
&\quad +\beta_1(x_t - \Omega_1 x_{t-1} - . - \Omega_m x_{t-m}) + (\tau_0 + \tau_1 x_t)\varepsilon_t
\end{aligned}
$$

The errors ν_t, ε_t are assumed to be independent of all x_{t-j}. $j \geq 0$.

The estimation of models where the states $x_t, x_{t-1}...x_{t-m}$ are not observed is usually carried out by using a Kalman filter. The main differences of the approach used by Hamilton and followed here relative to the standard Kalman filter are the non-linearities in the parameters and the fact that the states are assumed to follow a Markov process. An optimal nonlinear inference of the states is carried out by the bivariate extension to Hamilton's filter presented below.

4.1 The Filter.

We assume that both the variables included in the filter are governed by the scalar state variable. The filter involves the following five steps.

Step_1. Calculate the joint density of the m past states and the current state conditional on the information included in S_{t-1} ,D_{t-1} and all past values of S and D, S and D being the variables that are observed.

$$
\begin{aligned}
& p(x_t, x_{t-1}, ..., x_{t-m} | S_{t-1}, S_{t-2}, ., S_0, D_{t-1}, D_{t-2}, ., D_0) \\
= \; & p(x_t|x_{t-1}) p(x_{t-1}, x_{t-2}, ., x_{t-m} | S_{t-1}, S_{t-2}, ., S_0, D_{t-1}, D_{t-2}, ., D_0)
\end{aligned}
$$

$p(x_t|x_{t-1})$ is transition probability matrix of the states which are assumed to follow a Markov process. As in all the subsequent steps, the second term on the right-hand-side is known from the preceding step of the filter, in this case, $p(x_{t-1}, ..., x_{t-m}|S_{t-1}, ..., S_0, D_{t-1}, .., D_0)$ is known from the input to the filter, which in turn represents the result of the iteration at date $t-1$ (from step 5).

As in the univariate case, we require initial values for the parameters and initial conditions for the Markov process. The unconditional distribution $p(x_m, x_{m-1}, ., x_0)$ has been chosen for the first observation.

Step_2. Calculate the joint conditional distribution of S_t, D_t and $(x_t, x_{t-1}, ..., x_{t-m})$
$p(S_t, D_t, x_t, x_{t-1}, ., x_{t-m}|S_{t-1}, ., S_0, D_{t-1}, ., D_0)$
$= p(S_t, D_t|x_t, x_{t-1}, ., x_{t-m}, S_{t-1}, ., S_0, D_{t-1}, ., D_0) p(x_t, x_{t-1}, ., x_{t-m}|S_{t-1}, ., S_0, D_{t-1}, ., D_0)$
where we assume that

$$
\begin{aligned}
& p(S_t, D_t | x_t, x_{t-1}, ..., x_{t-m}, S_{t-1}, .., S_0, D_{t-1}, ..., D_0) \\
= \; & \frac{1}{2\pi|\Sigma|^{1/2}} \exp(-\frac{1}{2} u' \Sigma^{-1} u)
\end{aligned}
$$

$$u = \begin{bmatrix} (S_t - \alpha_1 x_t - \alpha_0) - \Phi_m(S_{t-m} - \alpha_1 x_{t-m} - \alpha_0) - \psi_1(D_{t-1} - \beta_1 x_{t-1} - \beta_0) \\ -. - \psi_m(D_{t-m} - \beta_1 x_{t-m} - \beta_0) \\ (D_t - \beta_1 x_t - \beta_0) - \varphi_m(S_{t-m} - \alpha_1 x_{t-m} - \alpha_0) - \Omega_1(D_{t-1} - \beta_1 x_{t-1} - \beta_0) \\ -. - \Omega_m(D_{t-m} - \beta_1 x_{t-m} - \beta_0) \end{bmatrix}$$

$$\Sigma_{x_0} = \begin{bmatrix} \omega_0^2 & c_0 \\ c_0 & \tau_0^2 \end{bmatrix} \qquad \Sigma_{x_0} = \begin{bmatrix} \omega_1^2 & c_1 \\ c_1 & \tau_1^2 \end{bmatrix}$$

Note that $p(S_t D_t | x_t, .., x_{t-m}, S_{t-1}, ..., S_0, D_{t-1}, ., D_0)$ involves $(x_t, ., x_{t-m})$ which is a vector which can take 2^{m+1} values.

Step_3. Marginalize the previous joint density with respect to the states which gives the conditional density from which the (conditional) likelihood function is calculated.

$$\begin{aligned} & p(S_t, D_t | S_{t-1}, ..., S_0, D_{t-1}, ..., D_0) \\ = & \sum_{x_t=0}^{1} \sum_{x_{t-1}=0}^{1} ... \sum_{x_{t-m}=0}^{1} p(S_t, D_t, x_t, ..., x_{t-m} | S_{t-1}, ..., S_0, D_{t-1}, ..., D_0) \end{aligned}$$

Step_4. Combining the results from steps 2 and 3, calculate the joint density of the state conditional on the observed current and past realizations of y.

$$\begin{aligned} & p(x_t, x_{t-1}, ..., x_{t-m} | S_t, ..., S_0, D_t, ..., D_0) \\ = & \frac{p(S_t, D_t, x_t, ..., x_{t-m} | S_{t-1}, ..., S_0, D_{t-1}, ..., D_0)}{p(S_t, D_t | S_{t-1}, ..., S_0, D_{t-1}, ..., D_0)} \end{aligned}$$

Step_5 The desired output is then obtained from

$$\begin{aligned} & p(x_t, x_{t-1}, ..., x_{t-m+1} | S_t, ..., S_0, D_t, ..., D_0) \\ = & \sum_{x_{t-m}=0}^{1} p(x_t, x_{t-1}, ..., x_{t-m} | S_t, ..., S_0, D_t, ..., D_0) \end{aligned}$$

The output of step 5 is used as an input to the filter in the next iteration. Estimates of the parameters are calculated as a by-product of the filter from step 3.

TESTING THE TERM STRUCTURE OF INTEREST RATES FROM A SWITCHING REGIME VAR

The process that drives the spread and the short-term interest rate difference is the VAR of equation described above in which D_t denotes the first difference of the three month rate, $R_{1t} - R_{1t-1}$ and S_t denotes the yield spread $R_{2t} - R_{1t}$.

The expectations hypothesis of the term structure of the interest rates can be written as

$$S_t = (1/2)E_t D_{t+1} + \theta + u_t$$

The restrictions imposed by the expectations model are presented below. Both an unrestricted and a restricted VAR can be estimated, and the restrictions tested using a likelihood ratio test.

Derivation of the restrictions in the regime-shifting VAR

We express the system in companion form:

$$\begin{bmatrix} S'_t \\ S'_{t-1} \\ S'_{t-2} \\ S'_{t-3} \\ D_t \\ D'_{t-1} \\ D'_{t-2} \\ D'_{t-3} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 & \psi_1 & \psi_2 & \psi_3 & \psi_4 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & \Omega_1 & \Omega_2 & \Omega_3 & \Omega_4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} S'_{t-1} \\ S'_{t-2} \\ S'_{t-3} \\ S'_{t-4} \\ D_{t-1} \\ D'_{t-2} \\ D'_{t-3} \\ D'_{t-4} \end{bmatrix} + \begin{bmatrix} (\omega_0 + \omega_1 x_t)\nu_t \\ 0 \\ 0 \\ 0 \\ (\tau_0 + \tau_1 x_t)\varepsilon_t \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{A3}$$

This allows us to express the expected values of the centered interest rate differences conditional on information at time $t-1$. We have proceeded as in Hamilton (1988), first considering forecasts under the assumption that the information set includes observation of the regime directly. This information set can be written as $I_t^* = \{S_t, S_{t-1}, ., D_t, D_{t-1}, ., x_t, x_{t-1}, .,\}$ This assumption will be relaxed below. Using this notation we can express the expected two periods ahead centered short-term interest rate difference conditional on information possessed by the agents at time $t-1$ as:

$$E[D'_{t+1}|I^*_{t-1}] = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \Delta^2 Z'_{t-1},$$

where

$$\Delta = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 & \psi_1 & \psi_2 & \psi_3 & \psi_4 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 & \Omega_1 & \Omega_2 & \Omega_3 & \Omega_4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix};$$

$$Z'_{t-1} = \begin{bmatrix} S_{t-1} - \alpha_0 - \alpha_1 x_{t-1} \\ S_{t-2} - \alpha_0 - \alpha_1 x_{t-2} \\ S_{t-3} - \alpha_0 - \alpha_1 x_{t-3} \\ S_{t-4} - \alpha_0 - \alpha_1 x_{t-4} \\ D_{t-1} - \beta_0 - \beta_1 x_{t-1} \\ D_{t-2} - \beta_0 - \beta_1 x_{t-2} \\ D_{t-3} - \beta_0 - \beta_1 x_{t-3} \\ D_{t-4} - \beta_0 - \beta_1 x_{t-4} \end{bmatrix}$$

The expected 2 period ahead first-difference of the short-term interest rate can be written as:

$$E[D_{t+1}|I^*_{t-1}] = \beta_0 + \beta_1 E(x_{t+1}|I^*_{t-1}) + E(D'_{t+1}|I^*_{t-1}),$$

and by further substitution

$$E[D_{t+1}|I^*_{t-1}] = \beta_0 + \beta_1 E(x_{t+1}|I^*_{t-1}) + \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \Delta^2 Z'_{t-1}$$

A similar formula can be derived for the spread:

$$E[S_t|I^*_{t-1}] = \alpha_0 + \alpha_1 E(x_t|I^*_{t-1}) + E(S'_t|I^*_{t-1})$$

and by further substitution

$$E[S_t|I^*_{t-1}] = \alpha_0 + \alpha_1 E(x_t|I^*_{t-1}) + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Delta Z'_{t-1}$$

Substituting the expected value of a 2-period-ahead two-state Markov process, the following expression for the expected short rate is obtained:

$$E[D_{t+1}|I^*_{t-1}] = \beta_0 + \beta_1[\rho + (x_{t-1} - \rho)\lambda^2] + \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \Delta^2 Z'_{t-1}$$

where $\rho = \frac{1-q}{2-p-q}$ and $\lambda = (p + q - 1)$

Finally, substituting the expected (difference of the) short-term interest rates and the expected spread, the term structure of interest rates relationship can be expressed as;

$$\begin{aligned} & \alpha_0 + \alpha_1[\rho + (x_{t-1} - \rho)\lambda] + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Delta Z'_{t-1} \qquad ((A4)) \\ = & \frac{1}{2}[\beta_0 + \beta_1[\rho + (x_{t-1} - \rho)\lambda^2] + \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \Delta^2 Z'_{t-1}] + \theta \end{aligned}$$

where

$$\Delta^2 = \left[\begin{array}{c|c} \text{A} & \text{B} \\ \hline \text{C} & \text{D} \end{array}\right] \qquad ((A5))$$

$$A = \begin{bmatrix} \Phi_1^2 + \Phi_2 + \varphi_1\psi_1 & \Phi_1\Phi_2 + \Phi_3 + \varphi_2\psi_1 & \Phi_1\Phi_3 + \Phi_4 + \varphi_3\psi_1 & \Phi_1\Phi_4 + \varphi_4\psi_1 \\ \Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} \psi_1\Phi_1 + \psi_2 + \psi_1\Omega_1 & \psi_1\Omega_2 + \psi_3 + \psi_2\Phi_1 & \Phi_1\psi_3 + \psi_4 + \Omega_3\psi_1 & \Phi_1\psi_4 + \Omega_4\psi_1 \\ \psi_1 & \psi_2 & \psi_3 & \psi_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} \varphi_1\Phi_1 + \varphi_2 + \varphi_1\Omega_1 & \varphi_2\Omega_1 + \varphi_3 + \varphi_1\Phi_2 & \Phi_3\varphi_1 + \varphi_4 + \Omega_1\varphi_3 & \Phi_4\varphi_1 + \Omega_1\varphi_4 \\ \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} \Omega_1^2 + \Omega_2 + \varphi_1\psi_1 & \Omega_1\Omega_2 + \Omega_3 + \varphi_1\psi_2 & \Omega_1\Omega_3 + \Omega_4 + \varphi_1\psi_3 & \Omega_1\Omega_4 + \varphi_1\psi_4 \\ \Omega_1 & \Omega_2 & \Omega_3 & \Omega_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Expression (A2) is linear in $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$. This means that to calculate the forecasts given information on I_{t-1} , we only have to replace $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ in expression (A4) with their conditional expectation given I_{t-1} .

The restrictions on the VAR follow immediately from equating the first row of the matrix Δ with the 5^{th} row of Δ^2 , and by equating the constants and coefficients of x_{t-1} on each side of the equation (A4).

The restrictions written out in full are:

$\Phi_1(2-\varphi_1) = \varphi_2 + \varphi_1\Omega_1$	$\psi_1(2-\varphi_1) = \Omega_2 + \Omega_1\Omega_1$
$\Phi_2(2-\varphi_1) = \varphi_3 + \varphi_2\Omega_1$	$\psi_2(2-\varphi_1) = \Omega_3 + \Omega_2\Omega_1$
$\Phi_3(2-\varphi_1) = \varphi_4 + \varphi_3\Omega_1$	$\psi_3(2-\varphi_1) = \Omega_4 + \Omega_3\Omega_1$
$\Phi_4(2-\varphi_1) = \varphi_4\Omega_1$	$\psi_4(2-\varphi_1) = \Omega_4\Omega_1$
$\alpha_1 = \frac{\beta_1\lambda}{2}$	

RESULTS OF THE BIVARIATE FILTER FOR SPREAD AND FIRST DIFFERENCE OF SHORT TERM RATES.

The filter was run for USA data for the period March1962-September1987. With no restrictions imposed, the results are as follows. The estimated unconditional means are:

$$\begin{array}{lll} \hat{mean}_S(x_t) = & .2479 & +0.8880x_t \\ & (3.6566) & (4.3586) \\ \hat{mean}_D(x_t) = & .0399 & +1.6210x_t \\ & (1.4719) & (5.4925) \end{array}$$

The estimated covariance matrices for state 0 and state 1 are:

$$\hat{\Sigma}_{x_0} = \begin{bmatrix} .3100 & .0309 \\ (6.3461) & (1.5993) \\ & .0885 \\ & (6.4605) \end{bmatrix} \qquad \hat{\Sigma}_{x_1} = \begin{bmatrix} 6.5448 & 1.9526 \\ (3.3852) & (4.9799) \\ & 1.0265 \\ & (2.3377) \end{bmatrix}$$

The probabilities of remaining in the same state as the previous period are:

$$\begin{array}{llll} \hat{p} = & .8373 & \hat{q} = & .9782 \\ & (5.2592) & & (64.3921) \end{array}$$

with expected duration of the states of

$$(1-\hat{q})^{-1} = 46.0630 \text{ months}$$
$$(1-\hat{p})^{-1} = 6.1482 \text{ months} .$$

The results of the centred vector autoregression are presented below.

$$
\begin{array}{rcccc}
S'_t = & -.0737S'_{t-1} & -.1259S'_{t-2} & +.3535S'_{t-3} & -.0389S'_{t-4} \\
 & (-.6738) & (-.5622) & (1.5599) & (-.1668) \\
 & +1.880D'_{t-1} & -.2927D'_{t-2} & +.1185D'_{t-3} & -.0617D'_{t-4} \\
 & (.9370) & (-1.4716) & (.5829) & (-.8388) \\
\hat{D}'_t = & .9206S'_{t-1} & -.2758S'_{t-2} & +.2757S'_{t-3} & +.1659S'_{t-4} \\
 & (16.6812) & (-2.5431) & (2.4383) & (1.3912) \\
 & +.3255D'_{t-1} & -.2250D'_{t-2} & -.0992D'_{t-3} & -.0092D'_{t-4} \\
 & (3.2131) & (-2.1276) & (-0.9309) & (-.2427)
\end{array}
$$

The results for the restricted VAR are as follows. The estimated unconditional means are:

$$
\begin{array}{rcc}
me\hat{a}n_S(x_t) = & .2559 & +0.7315x_t \\
 & (3.0395) & \\
me\hat{a}n_D(x_t) = & .0460 & +1.6946x_t \\
 & (.4510) & (6.3019)
\end{array}
$$

The estimated covariance matrix for state 0 and state 1 are:

$$
\hat{\Sigma}_{x_0} = \begin{bmatrix} .3171 & .0309 \\ (6.1881) & (1.6019) \\ & .0889 \\ & (6.7281) \end{bmatrix} \qquad \hat{\Sigma}_{x_1} = \begin{bmatrix} 6.6380 & 2.2090 \\ (6.1218) & (5.0504) \\ & 1.2778 \\ & (2.9433) \end{bmatrix}
$$

The probabilities of remaining in the same state as the previous period are:

$$
\begin{array}{rc}
\hat{p} = & .8835 \\
 & (11.0581) \\
\hat{q} = & .9798 \\
 & (67.7155)
\end{array}
$$

with expected duration of the states of

$$
\begin{array}{l}
(1-\hat{q})^{-1} = 49.6150 \text{ months} \\
(1-\hat{p})^{-1} = 8.5874 \text{ months} .
\end{array}
$$

The results of the centred vector autoregression are presented below.

$$
\begin{array}{rcccc}
\widehat{S}'_t = & .0059S'_{t-1} & +.1006S'_{t-2} & +.2345S'_{t-3} & +.0566S'_{t-4} \\
 & -0.0486D'_{t-1} & -.1541D'_{t-2} & +.0245D'_{t-3} & -.0343D'_{t-4} \\
\hat{D}'_t = & .9305S'_{t-1} & -.3063S'_{t-2} & +.2105S'_{t-3} & +.1801S'_{t-4} \\
 & (16.2174) & (-3.1903) & (2.0875) & (1.6931) \\
 & +.3360D'_{t-1} & -.1650D'_{t-2} & -.1094D'_{t-3} & +.0105D'_{t-4} \\
 & (3.6233) & (-1.8035) & (-1.0734) & (.2232)
\end{array}
$$

The likelihood ratio for testing the restrictions in the restricted VAR is asymptotically distributed under the null as Chi square with ten degrees of freedom since ten restrictions were imposed. The result is: $2(34.2325 - 31.6316) = 5.2018$. Thus the hypothesis that the restrictions implied by the term structure of interest rates is not rejected at the 5% level.[1]

The restricted estimates show that the mean change in the short rate was not significantly different from zero in state0, but was positive in state1, the point estimate of the mean rate of change being 1.7 percentage points per quarter. The variance covariance matrix differs between states. The variance of innovations in the spread is bigger by a factor of 21 in state 1 than in state0, and the variance of differences in the short rate by a factor of 13. The covariance between innovations is insignificant in state 0 and significantly different from zero in state 1.

On the basis of these estimates, we can calculate the probabilities of having been in each of the two states at any time, i.e., $prob(x_t = 1|S_t, S_{t-1}, S_{t-2}, ..., D_t, D_{t-1}, D_{t-2}, ...)$. These are illustrated in figures 6 and 7. The probabilities obtained from the unrestricted estimates are very similar to those obtained from the restricted estimates. These indicate a shift to the high variance regime only for the period 1979(4)-1982(3). These results broadly confirm Hamilton's finding of a regime change associated with the change in Federal Reserve operating procedures at that time.

6. MONTE CARLO ANALYSIS.

In order to shed some light on the usefulness of this methodology, this section reports the results of a small Monte Carlo investigation of the issues in question. (This is analogous to an exercise undertaken in a related context, stock prices, allowing for stochastic regime switching, by Cecchetti, Lam, and Mark(1990)). The data generating process in the Monte Carlo study is the restricted VAR with endogenous regime-switching of the previous section. This process is by construction consistent with the expectations model of the term structure. We then perform the traditional tests described in section 2 and calculate their small-sample properties.

In all simulations that follow, each Monte Carlo experiment is replicated 1000 times, using the GAUSS matrix programing language and its RNDN function to generate the pseudo-normal innovations.

In each replication, the initial values are the four first historical values of D and S (S and D between 1962 (2) and 1963 (2)). For x, the corresponding initial values are 0,0,0,0. The results are therefore conditional on these fixed initial values. We then generate a sequence of realizations of the Markov process for x, and innovations in each of the two equations in the VAR, generating a sample of 102 observations. These are used to generate series for the spread S and the difference in the short rate D. The following test statistics are then calculated:

[1] Note that the log lokelihoods quoted in the text omi a constant term equal in value to $-N\ln(2\pi)$, where N is the number of observations

(i) t-statistic for the null hypothesis $d_1 = 2$ in the Mankiw-Miron OLS test regression:

$$D_t = c_1 + d_1 S_{t-1} + \zeta_{1t}.$$

(ii) t-statistic for the null hypothesis $d_1 = 2$ in the previous test regression estimated by IV; the instrument set includes $S_{t-2}, S_{t-3}, S_{t-4}, D_{t-1}, D_{t-2}, D_{t-3}, D_{t-4}$, and a constant term.

(iii) t-statistic for the null hypothesis $d_2 = .5$ in the test regression:

$$S_t = c_2 + d_2 D_{t+1} + \zeta_{2t},$$

estimated by IV; the instrument set includes $D_t, D_{t-1}, D_{t-2}, D_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}$, and a constant term.

Table2 reports the estimated mean biases $(\hat{d}_1 - 2)$ and $(\hat{d}_2 - .5)$ together with their Monte Carlo standard error, as well as rejection frequencies of the expectations hypothesis under the estimation strategies (i), (ii) and (iii). All tests are two-sided, and are performed using 5% Gaussian critical values. Since 1000 replications per experiment are generated, the approximate 95% confidence interval for the test rejection frequencies for a nominal size of 5% is [3.6%,6.4%].

It is apparent from the rejection frequencies of the test in case (i) that inferences drawn from the Mankiw-Miron regression are wholly unreliable if the data is drawn from a switching regime-restricted VAR DGP The OLS estimator of the slope coefficient is severely biased, and the relevant t-test rejects the null hypothesis an impressive 100% of the time.

The use of an IV estimator in the same regression does not provide a substantial improvement: the estimate of d remains substantially biased, and the test statistic in case (ii) massively over-rejects the (true) null hypothesis. The empirical size of the test is 95.7% (93.8% using White's correction for heteroskedasticity) against a nominal size of 5%.

The test in case (iii) appears less bad than the other two in terms of the average size of the distortion. Nevertheless, its performance is still very poor under the assumed data generating process. The bias of the IV estimator of d is smaller in absolute size than the biases obtained in cases (i) and (ii), but as a proportion of the true value of the parameter it is as bad as the worst of them. The associated t-test rejects in 75.5% of replications.

The use of heteroskedasticity-consistent standard errors in case (iii) makes a small reduction in the frequency of rejection. Nevertheless, the test still rejects in 64.2% of replications.

	(i)	*(ii)*	*(iii)*
Rejection frequency (in %)	100	95.7	75.5
(using White´s correction)	100	93.8	64.2
Estimated mean bias	-.95	-.75	.22
(of the sample coef.)			
M.C. Standard error	(.0033)	(.0055)	(.0037)
Proportional bias	-.475	-.375	.44
Average std	(.05969)	(.1265)	(.07413)
Average White´s std	(.08335)	(.1636)	(.09893)

The results of the Monte Carlo experiments show that the values obtained for the slope coefficient in section (3) for the full sample are not inconsistent with the data having been generated from the restricted VAR with regime switching.

It may be useful to contrast these results with ones obtained in a very similar situation, except that the ingredient of regime switching is absent. A related paper (Driffill et al(1992)) uses a Monte Carlo study of artificial interest rate data, allowing as here for an error in the expectations model. It uses time-series processes with unit roots and near unit roots for the short rate. The sample size is 100. It finds that the frequency of rejection in case (i) above is 100%; in case (ii) it is around 50%; and in case (iii) it is around 7% (the exact frequency depending on the time-series process assumed for the short term interest rate). This suggests that, for the Mankiw-Miron regression (cases (i) and (ii) above), regime-switching is only partly responsible for the high rejection frequencies. Even though IV in the Mankiw-Miron regression is consistent in the absence of regime switching, it turns out to be biased in small samples. However, in case (iii) above, the small sample properties of the estimator appear to be satisfactory in the absence of regime-switching.

Only when the sample size is increased to 10000 observations (iii) gives a 5% frequency of rejections.