

Forecasting Markov-switching dynamic, conditionally heteroscedastic processes

James Davidson*
Cardiff University

Final Version, January 2004

Keywords: Forecasts, Markov-switching, ARFIMA, ARCH.

JEL Classification: C22

Abstract

Recursive formulae are derived for the multi-step point forecasts and forecast standard errors of Markov switching models with $\text{ARMA}(\infty, q)$ dynamics (including the fractionally integrated case) and conditional heteroscedasticity in $\text{ARCH}(\infty)$ form. Hamilton's dynamic models of switching mean and variance are also treated, in a slightly modified version of the analysis.

1 Introduction

Computing multi-step forecasts for nonlinear dynamic models is inherently difficult because there is no natural way to compute the conditional expectation of the future path of the process. In general, substituting the expected values of future shocks into the model equation will not achieve this. The only solution to the problem with a general application appears to be Monte Carlo simulation. It has been pointed out by a number of authors (e.g. Clements and Krolzig 1997, Krolzig 2002, Blix 1999) that although Markov-switching models are nonlinear, they have features that permit an analytical solution of the forecasts, at least in simple cases. However, these applications have dealt in practice with relatively simple models, finite order VARs with just one switching component, such as the mean of the process.

This note considers the general problem of multi-step forecasting with a Markov-switching dynamic regression model, possibly featuring conditional heteroscedasticity. We shall be interested in the point forecasts, but the calculation of forecast standard errors is of equal importance and proves to be the more computationally challenging problem. For simplicity a univariate model is treated. Thus, the framework envisaged can be written in the form

$$y_t = \alpha_{s_t} + \sum_{m=1}^{\infty} \phi_{mS_t} y_{t-m} + \sum_{m=1}^q \theta_{mS_t} u_{t-m} + u_t \quad (1.1)$$

where $u_t = h_t^{1/2} e_t$, with $e_t \sim i.i.d.(0, 1)$ and

$$h_t = \omega_{S_t} + \sum_{m=1}^{\infty} \beta_{mS_t} u_{t-m}^2 \quad (1.2)$$

*Email: davidsonje@cardiff.ac.uk. Research supported by the ESRC under award L138251025..

The reason for allowing an infinite order autoregressive structure in (1.1) is to include fractionally integrated (long memory) models such as the ARFIMA (see Granger and Joyeux 1980, Beran 1994, Baillie 1996). Similarly, equation (1.2) represents the so-called ARCH(∞) class, that includes the regular GARCH(p, q) model as well as cases such as FIGARCH and HYGARCH (see Baillie et. al 1996, Davidson 2004). Adding exogenous variables that can be assumed known in the forecast period, such as trends, is a straightforward extension of the results. The intercept α_{s_t} can effectively stand in for these, in the notation.

In this class of models, any of the coefficients can potentially be a random variable, switching between a discrete set of alternative values. Taking the case of α for illustration, we define

$$\alpha_{S_t} = \sum_{j=1}^M 1_j(t) \alpha_j$$

where

$$1_j(t) = \begin{cases} 1 & S_t = j \\ 0 & \text{otherwise} \end{cases}$$

and S_t denotes the regime prevailing at time t , one of M possible cases. The model is completed by postulating an exogenous Markov process that generates switches of regime with certain fixed probabilities $\{p_{ji}\}$ depending on the current state, such that $p_{ji} = \Pr(S_t = j | S_{t-1} = i)$. The probability that the observed process is in a particular state j at time t can then be calculated from the filtering (updating) equation

$$\Pr(S_t = j | \Psi_t) = \frac{f(y_t | S_t = j, \Psi_{t-1}) \Pr(S_t = j | \Psi_{t-1})}{\sum_{i=1}^M f(y_t | S_t = i, \Psi_{t-1}) \Pr(S_t = i | \Psi_{t-1})}. \quad (1.3)$$

where Ψ_t denotes information known at time t (specifically, $y_{t-j}, S_{t-j}, j \geq 0$) and

$$\Pr(S_t = j | \Psi_{t-1}) = \sum_{i=1}^M p_{ji} \Pr(S_{t-1} = i | \Psi_{t-1}). \quad (1.4)$$

For details of models of this class, see for example Hamilton (1994), Kim and Nelson (1999) and Gray (1996). An assumption is needed about the form of $f(\cdot | S_t = j, \Psi_{t-1})$, but it does not need to be Gaussian. In many financial data sets, for example, the Student t distribution with degrees of freedom > 2 could be a better choice. Another easily available extension is to have the transition probabilities p_{ji} depend on exogenous variables, instead of being constant parameters. Provided the endogenous mechanism driving the y_t process can be conditioned on these variables, this variant is accommodated directly.

2 Point Forecasts

In this setup, consider the problem of forecasting y_{T+K} for $K \geq 1$, given observations on the process up to date T only. We assume specifically that the object of interest is $E(y_{T+K} | \Psi_T)$. We abstract from the problem of unknown parameters, and assume the information available to the forecaster consists of $y_t, S_t, t \leq T$, and the values of the sets $\{\phi_{mj}, \theta_{mj}, \beta_{mj}, m \geq 1, \alpha_j, \omega_j\}$ for $j = 1, \dots, M$, and $p_{ji}, i, j = 1, \dots, M$. We can also treat the realized sample probabilities P_{T1}, \dots, P_{TM} as known, since they can be computed from the sample data using the recursion (1.3)+(1.4). To simplify notation, and with no loss of generality, we will set $T = 0$ in what follows.

Although the notation of (1.1) describes the evolution of observed y_t unambiguously in terms of past realized values of the variable, it does not identify the fact that the distributions of y_t

and u_t depend on the history of all the regime switches up to date, through dependence on y_{t-k} and u_{t-k} for $k = 1, 2, 3, \dots$. We can make this explicit, in the case of the k -step ahead realized value y_K , by writing (with the convention that a sum in which the lower limit exceeds the upper is zero) the recursions

$$\begin{aligned} y_K(j_1, \dots, j_K) &= \alpha_{j_K} + \sum_{m=1}^{K-1} \phi_{m,j_K} y_{K-m}(j_1, \dots, j_{K-m}) + \sum_{m=K}^{\infty} \phi_{m,j_K} y_{K-m} \\ &\quad + u_K(j_1, \dots, j_K) + \sum_{m=1}^{\min(q,K)-1} \theta_{m,j_K} u_{K-m}(j_1, \dots, j_{K-m}) \\ &\quad + \sum_{m=\min(q,K)}^q \theta_{m,j_K} u_{K-m} \end{aligned}$$

where $u_t(j_1, \dots, j_K) = e_t h_t^{1/2}(j_1, \dots, j_K)$ and

$$h_t(j_1, \dots, j_K) = \omega_{j_K} + \sum_{m=1}^{\infty} \beta_{m,j_K} u_{K-m}(j_1, \dots, j_{K-m})^2. \quad (2.1)$$

Of course, $y_{K-m}(j_1, \dots, j_{K-m})$ and $u_{K-m}(j_1, \dots, j_{K-m})$ for $0 < m < K$ are defined in the same manner, whereas y_t and u_t for $t \leq 0$ are observed, so we can condition on these values. In this framework, the distribution of interest is that of

$$y_K = \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M 1_{j_1}(1) \cdots 1_{j_K}(K) y_K(j_1, \dots, j_K). \quad (2.2)$$

In other words, we have to consider the M^K different contingent states of the world (sequences of regime switches) that can give rise to the realization y_K we wish to forecast. Let the point forecast be written for brevity $\hat{y}_K = E(y_K | \Psi_0)$. Our first proposition provides an easy recursion for computing \hat{y}_K involving only M terms at each step, the terms being the probability-weighted averages of the one-step contingent forecasts with \hat{y}_{K-m} replacing y_{K-m} .

Proposition 1

$$\hat{y}_K = \sum_{j_K=1}^M \hat{P}_{Kj_K} \left[\alpha_{j_K} + \sum_{m=1}^{K-1} \phi_{m,j_K} \hat{y}_{K-m} + \sum_{m=K}^{\infty} \phi_{m,j_K} y_{K-m} + \sum_{m=\min(q,K)}^q \theta_{m,j_K} u_{K-m} \right].$$

where $\hat{P}_{tj} = \Pr(S_t = j | \Psi_0)$ is generated from

$$\hat{P}_{tj} = \sum_{i=1}^M p_{ji} \hat{P}_{t-1,i}, \quad j = 1, \dots, M, \quad t = 1, 2, \dots, K$$

Proof. Note that $1_{j_t}(t)$ is a Ψ_{t-1} -measurable random variable distributed independently of $\{e_s, s > 0\}$ by assumption, and that $E(1_{j_t}(t) | \Psi_0) = \hat{P}_{tj_t}$. The stated recursion for \hat{P}_{tj} follows from (1.4). Letting $\hat{y}_t(j_1, \dots, j_t) = E(y_t(j_1, \dots, j_t) | \Psi_0)$ denote the contingent forecast that would be made if the future regime switches were known, and the only uncertainty stemmed from the shocks e_t , $t > 0$, the relation

$$\hat{y}_t = \sum_{j_1=1}^M \cdots \sum_{j_t=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{tj_t} \hat{y}_t(j_1, \dots, j_t) \quad (2.3)$$

follows from (2.2) by the law of iterated expectations. The contingent forecasts are defined by simple recursions

$$\begin{aligned} \hat{y}_K(j_1, \dots, j_K) &= \alpha_{j_K} + \sum_{m=1}^{K-1} \phi_{k, j_K} \hat{y}_{K-m}(j_1, \dots, j_{K-m}) \\ &\quad + \sum_{m=K}^{\infty} \phi_{m, j_K} y_{K-m} + \sum_{m=\min(q, K)}^q \theta_{m, j_K} u_{K-m}. \end{aligned} \quad (2.4)$$

Using (2.4) with (2.3), and the fact that $\sum_{j=1}^M \hat{P}_{mj} = 1$ for each m , we have

$$\begin{aligned} \hat{y}_K &= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} \left(\alpha_{j_K} + \sum_{m=1}^{K-1} \phi_{m, j_K} \hat{y}_{K-m}(j_1, \dots, j_{K-m}) + \sum_{m=K}^{\infty} \phi_{m, j_K} y_{K-m} \right. \\ &\quad \left. + \sum_{m=\min(q, K)}^q \theta_{m, j_K} u_{K-m} \right) \\ &= \sum_{j_K=1}^M \hat{P}_{Kj_K} \left[\alpha_{j_K} + \sum_{m=1}^{K-1} \left(\phi_{m, j_K} \sum_{j_1=1}^M \cdots \sum_{j_{K-m}=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{mj_m} \hat{y}_{K-m}(j_1, \dots, j_{K-m}) \right) \right. \\ &\quad \left. + \sum_{m=K}^{\infty} \phi_{m, j_K} y_{K-m} + \sum_{m=\min(q, K)}^q \theta_{m, j_K} u_{K-m} \right] \end{aligned}$$

In view of the definition of \hat{y}_t , this expression simplifies to give our stated result. ■

3 Standard Errors of Forecast

Next, consider the standard error of the forecast. The uncertainty of the forecast has two sources, the stochastic driving process e_t , and the uncertainty about which regimes are occupied during the forecast period, and therefore, which model(s) should be used to construct the prediction. The derivation therefore falls broadly into two parts. First, define

$$S_K^2(j_1, \dots, j_K) = E[f_K(j_1, \dots, j_K)^2 | \Psi_0] \quad (3.1)$$

where

$$f_K(j_1, \dots, j_K) = y_K(j_1, \dots, j_K) - \hat{y}_K(j_1, \dots, j_K) \quad (3.2)$$

represents the forecast error that would be made if it were known that the process will pass through the sequence of regimes $\{j_1, \dots, j_K\}$, and the only uncertainty is the future values of e_t .

Proposition 2

$$\begin{aligned} &E[(y_K - \hat{y}_K)^2 | \Psi_0] \\ &= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} (S_K^2(j_1, \dots, j_K) + \hat{y}_K(j_1, \dots, j_K)^2) - \hat{y}_K^2. \end{aligned} \quad (3.3)$$

where \hat{y}_K is defined by (2.3).

Proof. Observe from (2.2), (2.3) and (3.2) that

$$\begin{aligned}
y_K - \hat{y}_K &= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M 1_{j_1}(1) \cdots 1_{j_K}(K) (f_K(j_1, \dots, j_K) \\
&\quad + \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M (1_{j_1}(1) \cdots 1_{j_K}(K) - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K}) \hat{y}_K(j_1, \dots, j_K)). \quad (3.4)
\end{aligned}$$

The realizations of the random variable

$$\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M 1_{j_1}(1) \cdots 1_{j_K}(K) f_K(j_1, \dots, j_K)$$

contain exactly one non-zero term, the one for which all the indicators are unity. Squaring this expression, note that the cross-product terms vanish identically since they all involve two indicators of the same date whose product is zero. Hence

$$\begin{aligned}
E \left[\left(\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M 1_{j_1}(1) \cdots 1_{j_K}(K) f_K(j_1, \dots, j_K) \right)^2 \middle| \Psi_0 \right] \\
= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} E[f_K(j_1, \dots, j_K)^2 | \Psi_0].
\end{aligned}$$

Also note, applying the law of iterated expectations and using $E_{t-1} 1_{j_i}(t) = \hat{P}_{tj_i}$ a.s. for $t = 1, \dots, K$, that

$$E[1_{j_1}(1) \cdots 1_{j_K}(K) - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} | \Psi_0] = 0 \text{ a.s.} \quad (3.5)$$

and

$$E[(1_{j_1}(1) \cdots 1_{j_K}(K) - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K})^2 | \Psi_0] = \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} (1 - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K}) \text{ a.s.}$$

whereas for $(i_1, \dots, i_K) \neq (j_1, \dots, j_K)$,

$$\begin{aligned}
E[(1_{j_1}(1) \cdots 1_{j_K}(K) - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K})(1_{i_1}(1) \cdots 1_{i_K}(K) - \hat{P}_{1i_1} \cdots \hat{P}_{Ki_K}) | \Psi_0] \\
= -\hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} \hat{P}_{1i_1} \cdots \hat{P}_{Ki_K} \text{ a.s.}
\end{aligned}$$

This establishes that

$$\begin{aligned}
E \left[\left(\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M (1_{j_1}(1) \cdots 1_{j_K}(K) - \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K}) \hat{y}_K(j_1, \dots, j_K) \right)^2 \middle| \Psi_0 \right] \\
= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} \hat{y}_K(j_1, \dots, j_K)^2 - \\
\left(\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} \hat{y}_K(j_1, \dots, j_K) \right)^2
\end{aligned}$$

where the second right-hand-side term is just \hat{y}_K^2 .

It remains to show that the cross-product of the terms in (3.4) vanishes almost surely in the conditional expectation. We know that the first of these terms has a mean of zero a.s., since their

sum does by construction, and also the second term, recalling $\hat{y}_K(j_1, \dots, j_K)$ is Ψ_0 -measurable. By elementary substitutions,

$$\begin{aligned}
f_K(j_1, \dots, j_K) &= \sum_{m=1}^{K-1} \phi_{m,j_K} f_{K-m}(j_1, \dots, j_{K-m}) + u_K(j_1, \dots, j_K) \\
&\quad + \sum_{m=1}^{\min(q,K)-1} \theta_{m,j_K} u_{K-m}(j_1, \dots, j_{K-m}) \\
&= u_K(j_1, \dots, j_K) + \sum_{m=1}^{K-1} \gamma_m(j_{K-m+1}, \dots, j_K) u_{K-m}(j_1, \dots, j_{K-m}). \tag{3.6}
\end{aligned}$$

where $\gamma_1(j_K) = \phi_{1,j_K} + \theta_{1,j_K}$, $\gamma_2(j_{K-1}, j_K) = (\phi_{1,j_K} + \theta_{1,j_K})\phi_{1,j_{K-1}} + \phi_{2,j_K} + \theta_{2,j_K}$, and so forth. Therefore, this contingent forecast error is just a linear combination of zero mean uncorrelated shocks. Since the regime-switching mechanism is exogenous, shocks and indicators are distributed independently and

$$E[1_{j_1}(1) \cdots 1_{j_K}(K) f_K(j_1, \dots, j_K) | \Psi_0] = 0 \text{ a.s.} \tag{3.7}$$

By a similar argument, the cross-product terms

$$E[1_{j_1}(1) \cdots 1_{j_K}(K) (f_K(j_1, \dots, j_K) (1_{i_1}(1) \cdots 1_{i_K}(K) - \hat{P}_{1i_1} \cdots \hat{P}_{Ki_K}) \hat{y}_K(i_1, \dots, i_K) | \Psi_0)]$$

are equal to zero a.s. for all pairs $\{j_1, \dots, j_K\}, \{i_1, \dots, i_K\}$. In the case $\{j_1, \dots, j_K\} = \{i_1, \dots, i_K\}$ this holds because

$$(1_{j_1}(1) \cdots 1_{j_K}(K))^2 = 1_{j_1}(1) \cdots 1_{j_K}(K)$$

and (3.7), and the fact that $\hat{P}_{1i_1} \cdots \hat{P}_{Ki_K} \hat{y}_K(i_1, \dots, i_K)$ is Ψ_0 -measurable. In the cases $\{j_1, \dots, j_K\} \neq \{i_1, \dots, i_K\}$ note that the product

$$1_{j_1}(1) \cdots 1_{j_K}(K) 1_{i_1}(1) \cdots 1_{i_K}(K) = 0$$

identically, and the conclusion follows as before, from (3.7) and Ψ_0 -measurability of

$$\hat{P}_{1i_1} \cdots \hat{P}_{Ki_K} \hat{y}_K(i_1, \dots, i_K).$$

This completes the proof. ■

The next step is to evaluate the $S_K^2(j_1, \dots, j_K)$ or, more precisely, to construct a recursive formula for the first block of terms of (3.3). This is as follows.

Proposition 3

$$\begin{aligned}
&\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} S_K^2(j_1, \dots, j_K) \\
&= \hat{h}_K + \sum_{m=1}^{K-1} \left(\sum_{j_{K-m+1}=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{K-m+1, j_{K-m+1}} \cdots \hat{P}_{Kj_K} \right. \\
&\quad \left. \times \gamma_m(j_{K-m+1}, \dots, j_K)^2 \right) \hat{h}_{K-m} \tag{3.8}
\end{aligned}$$

where $\hat{h}_t = E(h_t | \Psi_0)$ and $h_t = \sum_{j_1=1}^M \cdots \sum_{j_t=1}^M 1_{j_1}(1) \cdots 1_{j_t}(t) h_t(j_1, \dots, j_t)$ is the conditional variance of the disturbance realized at date t .

Proof. Since the u_t are uncorrelated and $u_t^2(j_1, \dots, j_t) = e_t^2 h_t(j_1, \dots, j_t)$, it follows from (3.6) that

$$S_K^2(j_1, \dots, j_K) = \hat{h}_K(j_1, \dots, j_K) + \sum_{m=1}^{K-1} \gamma_m(j_{K-m+1}, \dots, j_K)^2 \hat{h}_{K-m}(j_1, \dots, j_{K-m}).$$

Therefore,

$$\begin{aligned} & \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} S_K^2(j_1, \dots, j_K) \\ &= \sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} \hat{h}_K(j_1, \dots, j_t) \\ & \quad + \sum_{m=1}^{K-1} \left(\sum_{j_{K-m+1}=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{K-m+1, j_{K-m+1}} \cdots \hat{P}_{Kj_K} \gamma_m(j_{K-m+1}, \dots, j_K)^2 \right. \\ & \quad \left. \times \sum_{j_1=1}^M \cdots \sum_{j_{K-m}=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{K-m, j_{K-m}} \hat{h}_{K-m}(j_1, \dots, j_{K-m}) \right) \end{aligned} \quad (3.9)$$

However, note that

$$\sum_{j_1=1}^M \cdots \sum_{j_t=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{tj_t} \hat{h}_t(j_1, \dots, j_t) = E(u_t^2 | \Psi_0)$$

so that (3.9) acquires the more attractive form given in the proposition. ■

It remains to compute the sequence \hat{h}_t , which is provided by the following recursion.

Proposition 4

$$\hat{h}_t = \sum_{j_t=1}^M \hat{P}_{Kj_K} \left(\omega_{j_t} + \sum_{m=1}^{t-1} \beta_{mj_t} \hat{h}_{t-m} + \sum_{m=t}^{\infty} \beta_{mj_t} u_{t-m}^2 \right), \quad t = 1, \dots, K. \quad (3.10)$$

Proof. For $t \geq 1$,

$$\hat{h}_t = \sum_{j_1=1}^M \cdots \sum_{j_t=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{tj_t} h_t(j_1, \dots, j_t)$$

where

$$h_t(j_1, \dots, j_t) = \omega_{j_t} + \sum_{m=1}^{t-1} \beta_{mj_t} u_{t-m}^2(j_1, \dots, j_{t-m}) + \sum_{m=t}^{\infty} \beta_{mj_t} u_{t-m}^2.$$

The proposition follows by substitution, and the fact that $\sum_{j_t=1}^M \hat{P}_{tj_t} = 1$ for each $K-m < t < K$ and $1 \leq m < K$. ■

To compute

$$\sum_{j_1=1}^M \cdots \sum_{j_K=1}^M \hat{P}_{1j_1} \cdots \hat{P}_{Kj_K} S_K^2(j_1, \dots, j_K)$$

therefore requires a simple recursion based on (3.8) and (3.10).

The last result has made use of the fact that in the ARCH(∞) class of models, the conditional variance is linear in the squares of the process. Therefore, it does not generalize to other conditional heteroscedasticity models, in particular the EGARCH(∞) class in which $\log h_t$ is linear in

the lags of a function $g(e_t) = |e_t| + ae_t$ (say). We can construct $E(\log h_t|\Psi_0)$ similarly to (3.10), substituting $E[|e_t| + ae_t|\Psi_0] = E|e_t|$. By Jensen's inequality we can then say, at best, that the formula based on substitution of $\exp\{E(\log h_t|\Psi_0)\}$ for $E(h_t|\Psi_0)$ in (3.8) yields a lower bound on the standard error. See Nelson (1991) Appendix 1 for derivation of the true moments, which depend on the distribution of e_t .

4 Hamilton's Model

The models of Hamilton (1989) and Hamilton and Susmel (1994) can be treated in the same manner as the foregoing after a minor generalization. They fall into the AR(p)-ARCH(p) class,

$$y_t = \alpha_{s_t} + \sum_{m=1}^p \phi_m (y_{t-m} - \alpha_{s_{t-m}}) + u_t$$

$$h_t = \omega_{s_t} \left(1 + \sum_{m=1}^p \beta_m (u_{t-m}^2 / \omega_{s_{t-m}}) \right)$$

with the special feature that, in effect, the the intercept depends on the states of the last p periods, and the coefficients of u_{t-m}^2 depend similarly on the states at date $t - m$. A 'regime' is therefore defined as the configuration

$$\{S_t = j^0, S_{t-1} = j^1, \dots, S_{t-p} = j^p\}$$

for $j^m \in (1, \dots, M)$, with a total of M^{p+1} possibilities. However, the same analysis goes through as before, with a modified updating rule for filter probabilities. We omit the details of the filtering algorithm (see Hamilton 1989, and also Kim and Nelson 1999, Chapter 4, for a simplified account), apart from pointing out that (1.4) is replaced by

$$\begin{aligned} & \Pr(S_t = j^0, S_{t-1} = j^1, \dots, S_{t-p} = j^p | \Psi_{t-1}) \\ &= p_{j^0 j^1} \sum_{i^{p+1}=1}^M \Pr(S_{t-1} = j^1, S_{t-2} = j^2, \dots, S_{t-p-1} = i^{p+1} | \Psi_{t-1}). \end{aligned} \quad (4.1)$$

Let the possible states at date t be indexed by $j_t^m = 1, \dots, M$ for $m = 0, \dots, p$, and so let $\hat{P}_t(j_K^0, \dots, j_{K-p}^p)$ denote the associated projected probabilities. Then the counterparts of Propositions 1-4, which hold by just the same arguments with appropriate changes of notation, can be stated as follows.

Proposition 5

$$\begin{aligned} \hat{y}_K = & \underbrace{\sum_{j_K^0=1}^M \cdots \sum_{j_{K-p}^p=1}^M}_{M^{p+1}\text{-fold sum}} \hat{P}_K(j_K^0, \dots, j_{K-p}^p) \left(\alpha_{j_K^0} - \sum_{m=1}^p \phi_m \alpha_{j_{K-m}^m} \right) \\ & + \sum_{m=1}^{\min(p, K-1)} \phi_m \hat{y}_{K-m} + \sum_{m=\min(K-1, p)+1}^p \phi_m y_{K-m}. \end{aligned}$$

where

$$\hat{P}_t(j_t^0, \dots, j_{t-p}^p) = p_{j^0 j^1} \sum_{i^{p+1}=1}^M \hat{P}_{t-1}(j_{t-1}^1, \dots, j_{t-p}^p, i_{t-p-1}^{p+1})$$

for $j_t^m = 1, \dots, M$, $m = 0, \dots, p$, and $t = 1, \dots, K$.

Proposition 6

$$E[(y_K - \hat{y}_K)^2 | \Psi_0] = \underbrace{\sum_{j_1^0=1}^M \cdots \sum_{j_K^p=1}^M}_{M^{K(p+1)\text{-fold sum}}} \hat{P}_1(j_1^0, \dots, j_1^p) \cdots \hat{P}_K(j_K^0, \dots, j_K^p) \\ \times (S_K^2(j_1^0, \dots, j_K^p) + \hat{y}_K(j_1^0, \dots, j_K^p)^2) - \hat{y}_K^2.$$

where

$$\sum_{j_1^0=1}^M \cdots \sum_{j_K^p=1}^M \hat{P}_1(j_1^0, \dots, j_1^p) \cdots \hat{P}_K(j_K^0, \dots, j_K^p) S_K^2(j_1^0, \dots, j_K^p) \\ = \hat{h}_K + \sum_{m=1}^{K-1} \gamma_m^2 \hat{h}_{K-m}$$

where $\gamma_1 = \phi_1$, $\gamma_2 = \phi_1^2 + \phi_2$, etc. and

$$\hat{h}_t = \underbrace{\sum_{j_t^0=1}^M \cdots \sum_{j_{t-p}^p=1}^M}_{M^{p+1}\text{-fold sum}} \hat{P}_K(j_t^0, \dots, j_{t-p}^p) \omega_{j_t^0} \left(1 + \sum_{m=1}^{\min(p, t-1)} \beta_m \hat{h}_{t-m} / \omega_{j_{t-m}^m} \right) \\ + \sum_{m=\min(p, t-1)+1}^p \beta_m u_{t-m}^2 / \omega_{j_{t-m}^m},$$

for $t = 1, \dots, K$.

While there appears to be a simplification in the variance formula, thanks to the fact that the autoregressive coefficients are non-switching in this model, note that the $M^{K(p+1)}$ -fold sum is not avoidable in respect of the terms $\hat{y}_K(j_1^0, \dots, j_K^p)^2$.

5 Concluding Remarks

The Ox econometrics package Time Series Modelling 3.2 (Davidson 2003b, Doornik 1999) implements these formulae for the class of models treated here. Their main advantage over Monte Carlo methods, apart from being exact, should be their relative speed of evaluation. For the point forecasts this is certainly true, since K steps involve evaluating only MK ordinary one-step forecasts, or KM^{p+1} forecasts in Hamilton's model. However, the expression (3.3) contains of the order of M^K terms, which in the counterpart in Hamilton's model, becomes $M^{K(p+1)}$. Even with $M = 2$, in a model with several lags the computing time involved could turn out to be non-trivial for large K . A practical short-cut would to extrapolate the sequence beyond the point where it converges to a stable limit as K increases, provided this happens rapidly enough. Noting that three different dynamic processes (autoregression, ARCH and Markov chain) are potentially involved in this convergence, some possibly hyperbolic, this device will not serve in every case. However, standard error bands are presumably more useful in the first few steps than at long range, so this will be an acceptable limitation in most cases.

References

- Baillie, R. T. (1996) Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5–59
- Baillie, R. T., T. Bollerslev and H. O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 74, 3-30
- Beran, J (1994) *Statistics for Long Memory Processes*, Chapman and Hall
- Blix, M. (1999) Forecasting Swedish inflation with a Markov-switching VAR Working Paper No. 76. Central Bank of Sweden.
- Clements, M.P. and H.-M. Krolzig (1998) A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP *Econometrics Journal* 1, C47–C75.
- Davidson, J. (2004) Moment and memory properties of linear conditional heteroscedasticity processes, and a new model, *Journal of Business and Economics Statistics* 22 (1), pp 16-29.
- Davidson, J. (2003b) Time Series Modelling 3.25, <http://www.cf.ac.uk/carbs/econ/davidsonje/tsmod32.htm>
- Doornik, J. A. (1999) *Ox: An Object-oriented Matrix Programming Language*, Timberlake Consultants Ltd.
- Granger, C.W.J. and Roselyne Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 1, 15-29.
- Gray, S. F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27-62
- Hamilton, J. D. (1989) A new approach to the economics analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357-384.
- Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press.
- Hamilton, J. D. and R. Susmel (1994) Autoregressive conditional heteroscedasticity and changes in regime, *Journal of Econometrics* 64, 307–333.
- Kim, C.-J. and C. R. Nelson (1999) *State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches, with Applications*. MIT Press
- Krolzig, H.-M. (2000) Predicting Markov-Switching Vector Autoregressive Processes Working Paper, Department of Economics, University of Oxford.
- Nelson, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.