# Markov Switching GARCH Models: Filtering, Approximations and Duality

Monica Billio and Maddalena Cavicchioli

**Abstract** This paper is devoted to show duality in the estimation of Markov Switching (MS) GARCH processes. It is well-known that MS GARCH models suffer of path dependence which makes the estimation step unfeasible with usual Maximum Likelihood procedure. However, by rewriting the model in a suitable state space representation, we are able to give a unique framework to reconcile the estimation obtained by filtering procedure with that coming from some auxiliary models proposed in the literature. Estimation on short-term interest rates shows the feasibility of the proposed approach.

## 1 Introduction

Time varying volatility is one of the main property of many financial time series. Moreover, describing and, where possible, forecasting volatility is a key aspect in financial economics and econometrics. A popular class of models which describe time-varying volatility are Generalized Autoregressive Conditional Heteroschedasticity (GARCH) models. GARCH models ([8], [22], [21]) describe the variance as a linear function of the squares of past observations, so that one type of shock alone drives both the series itself and its volatility. One potential source of misspecification derives from the fact that structural forms of conditional means and variances are relatively inflexible and held fixed throughout the sample period. In this sense, they are called *single-regime* models since a single structure for the conditional mean and variance is assumed.

---

Monica Billio
Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
e-mail: billio@unive.it

Maddalena Cavicchioli
Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy
e-mail: maddalena.cavicchioli@univr.it

To allow more flexibility, the assumption of a single regime could be relaxed in favour of a *regime-switching* model. The coefficients of this model are different in each regime to account for the possibility that the economic mechanism generating the financial series undergoes a finite number of changes over the sample period. These coefficients are unknown and must be estimated, and, although the regimes are never observed, probabilistic statements can be made about the relative likelihood of their occurrence, conditional on an information set. A well-known problem to face when dealing with the estimation of Markov Switching (MS) GARCH models is the path dependence. [9] and [16] have argued that MS GARCH models are essentially intractable and impossible to estimate since the conditional variance depends on the entire path history of the data. That is, the distribution at time $t$, conditional on the current state and on available information, is directly dependent on the current state but also indirectly dependent on all past states due to the path dependence inherent in MS GARCH models. This is because the conditional variance at time $t$ depends upon the conditional variance at time $t-1$, which depends upon the regime at time $t-1$ and on the conditional variance at time $t-2$, and so on. Hence, the conditional variance at time $t$ depends on the entire sequence of regimes up to time $t$.

Some methods are proposed in the literature to overcome the problem of path dependence present in MS GARCH. The trick is mainly found in adopting different specifications of the original MS GARCH model. Some authors propose Quasi Maximum Likelihood (QML) procedures of a model which allows similar effects of the original one. Models which elude in this way the path dependence problem are proposed by [14], [10] and [18], among others, and are known as collapsing procedures. [14] proposes a model in which path dependence is removed by aggregating the conditional variances from the regimes at each step. This aggregated conditional variance (conditional on available information, but aggregated over the regimes) is then all that is required to compute the conditional variance at the next step. The same starting idea is used in [10], with a slightly different approach. The author extends the information set including also current information on the considered series. Furthermore, [18] puts further this idea. Particularly, when integrating out the unobserved regimes, all the available information is used, whereas [14] uses only part of it. Another method to deal with MS GARCH models has been proposed by [15] for which the variance is disaggregated in independent processes; this is a simple generalization of the GARCH process to a multi-regime setting. Furthermore, Bayesian approaches based on Markov Chain Monte Carlo Gibbs technique for estimating MS GARCH can be found in [3], [4], [17] or [5]. Other works based on both Monte Carlo methods combined with expectation-maximization algorithm and importance sampling to evaluate Maximum Likelihood (ML) estimators are conducted by [1], [6], and [7]. Finally, a recent paper by [2] proposes estimation of MS GARCH models with a deterministic particle filter.

The contribution of our paper is to give a unique framework to reconcile MS GARCH estimation obtained by the above auxiliary models from one side, and a filtering algorithm from the other. This relationship provides the missing link to justify the validity of approximations in estimating MS GARCH models. The use

of filtering is a flexible approach and it allows the estimation of a broad class of models that can be put in a switching state space form. However, to make the filter operable, at each iteration we need to collapse $M^2$ posteriors (where $M$ is the number of switching regimes) in $M$ of it, employing an approximation as suggested by [19]. Then, QML estimation of the model recovers the unknown parameters. Our algorithm is readily programmable and with a very limited computational cost. An empirical application shows the feasibility of this approach.

The paper is structured as follows. Section 2 introduces the MS GARCH model of interest and reviews the main auxiliary models proposed in the literature to overpass the path dependence problem. In Section 3 we present a filtering algorithm for MS GARCH models which serves to prove our duality results. In Section 4 we compare estimation of the parameters using different approximations in the proposed filters for financial data. Section 5 concludes. Finally, derivations of some formulae are given in the Appendix.

## 2 Markov Switching GARCH and its auxiliary models

Let $\varepsilon_t$ be the observed univariate time series variable (as for instance, returns on a financial asset) centered on its mean. The univariate MS GARCH(1,1) model is defined as

$$
\begin{cases}
\varepsilon_t = \sigma_t(\Psi_{t-1}, s_t) u_t \\
\sigma_t^2(\Psi_{t-1}, s_t) = \omega_{s_t} + \alpha_{s_t} \varepsilon_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-1})
\end{cases}
\tag{1}
$$

where $u_t \sim IID(0,1)$, $\omega_{s_t} > 0$, $\alpha_{s_t}, \beta_{s_t} \geq 0$. The state $s_t$ is a discrete, unobserved variable following a first order Markov chain with $M$ regimes and (time invariant) transition probabilities $\pi_{ij} = p(s_t = j | s_{t-1} = i)$, where $\sum_{j=1}^{M} \pi_{ij} = 1$, for every $i = 1, \ldots, M$. We assume that $(s_t)$ is independent of $(u_t)$. For necessary and sufficient stationarity conditions related with MS GARCH$(p, q)$ models, we refer to [12], Theorems 1 and 2. Consistency of maximum likelihood estimates, $L^2$ structure and inference for univariate MS GARCH models have been investigated by [11], [12] and [3]. Here the common approach to eliminate path dependence is to replace the lagged conditional variance derived from the original MS GARCH model with a proxy. Various authors have proposed different auxiliary models which differ only by the content of the information used to define such a proxy. In general, different auxiliary models can be obtained by approximating the conditional variance of process in (1) with

$$
\sigma_t^2(\Psi_{t-1}, s_t) = \omega_{s_t} + \alpha_{s_t} {}^{(SP)}\varepsilon_{t-1}^2 + \beta_{s_t} {}^{(SP)}\sigma_{t-1}^2.
\tag{2}
$$

In the literature there are different specifications (in short, SP) of ${}^{(SP)}\varepsilon_{t-1}^2$ and ${}^{(SP)}\sigma_{t-1}^2$ which in turn define different approximations of the original process. These collapsing procedures are illustrated below. First, we introduce some concepts and notations: $p(s_t = j | \Psi_{t-1}) = p_{j,t|t-1}$ are prediction probabilities; $p(s_t = j | \Psi_t) = p_{j,t|t}$

are filtered probabilities and from these we can compute augmented filtered probabilities as

$$p(s_{t-1} = i|s_t = j, \Psi_{t-1}) = \frac{\pi_{ij}\; p_{i,t-1|t-1}}{p_{j,t|t-1}} = p_{i,t-1|t,t-1}.$$

Here $\Psi_t$ denotes the information set of observations available up to time $t-1$. Note that the filtering algorithm computes $p_{t|t-1,t} = p(s_t|s_{t-1}, \Psi_t)$ in terms of $p_{t|t-1,t-1}$ and the conditional density of $\varepsilon_t$ which depends on the current regime $s_t$ and all past regimes, i.e, $f(\varepsilon_t|s_1, \dots, s_t, \Psi_{t-1})$. Computation details are shown in Appendix A1.

*2a. Gray's Model.* The first attempt to eliminate the path dependence is proposed by [14]. He approximates the original model by replacing the lagged conditional variance $\sigma_{t-1}^2$ with a proxy $^{(G)}\sigma_{t-1}^2$ as follows:

$$
\begin{aligned}
^{(G)}\sigma_{t-1}^2 &= E[\sigma_{t-1}^2(\Psi_{t-2}, s_{t-1})|\Psi_{t-2}] \\
&= \sum_{i=1}^{M} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-1} = i)\; p(s_{t-1} = i|\Psi_{t-2}) = \sum_{i=1}^{M} {}^{(G)}\sigma_{i,t-1|t-2}^2\; p_{i,t-1|t-2}
\end{aligned}
\tag{3}
$$

where, according to the model, $^{(G)}\sigma_{t-1|t-2}^2$ turns out to be a function of $\Psi_{t-2}$ and $s_{t-1} = i$. Note that the model originally proposed by Gray is not centered as in our case, but this can always be assumed without loss of generality.

*2b. Dueker's Model.* In the previous approximation, the information coming from $\varepsilon_{t-1}$ is not used. [10] proposes to change the conditioning scheme including $\varepsilon_{t-1}$ while assuming that $\sigma_{t-1}^2$ is a function of $\Psi_{t-2}$ and $s_{t-2}$. Hence

$$
\begin{aligned}
^{(D)}\sigma_{t-1}^2 &= E[\sigma_{t-1}^2(\Psi_{t-2}, s_{t-2})|\Psi_{t-1}] \\
&= \sum_{k=1}^{M} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-2} = k)\; p(s_{t-2} = k|\Psi_{t-1}) = \sum_{k=1}^{M} {}^{(D)}\sigma_{k,t-1|t-2}^2\; p_{k,t-2|t-1}
\end{aligned}
\tag{4}
$$

so that $^{(D)}\sigma_{t-1|t-2}^2$ is a function of $\Psi_{t-2}$ and $s_{t-2} = k$, and $p_{k,t-2|t-1}$ is one-period ahead smoothed probability which, shifting one period, can be computed as

$$p_{i,t-1|t} = p(s_{t-1} = i|\Psi_t) = p_{i,t-1|t-1} \sum_{j=1}^{M} \frac{\pi_{ij}\; p_{j,t|t}}{p_{j,t|t-1}}.$$

*2c. Simplified Klaassen's Model.* The approximation proposed by [18] is similar to that from [10] but it assumes that $\sigma_{t-1}^2$ is a function of $\Psi_{t-2}$ and $s_{t-1}$. So it results computationally simpler. In fact, we have

$$
\begin{aligned}
^{(SK)}\sigma_{t-1}^2 &= E[\sigma_{t-1}^2(\Psi_{t-2}, s_{t-1})|\Psi_{t-1}] \\
&= \sum_{i=1}^{M} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-1} = i)\; p(s_{t-1} = i|\Psi_{t-1}) = \sum_{i=1}^{M} {}^{(SK)}\sigma_{i,t-1|t-2}^2\; p_{i,t-1|t-1}.
\end{aligned}
\tag{5}
$$

Then from the considered model, $^{(SK)}\sigma^2_{t-1|t-2}$ results to be a function of $\Psi_{t-2}$ and $s_{t-1} = i$.

*2d. Klaassen's Model.* Finally, [18] generalizes the previous auxiliary model including in the conditioning set the information also coming from the current regime $s_t$. So $\sigma^2_{t-1}$ turns out to be approximated as

$$^{(K)}\sigma^2_{t-1} = E[\sigma^2_{t-1}(\Psi_{t-2}, s_{t-1})|\Psi_{t-1}, s_t = j]$$

$$= \sum_{i=1}^{M} \sigma^2_{t-1}(\Psi_{t-2}, s_{t-1} = i) \, p(s_{t-1} = i|s_t = j, \Psi_{t-1}) = \sum_{i=1}^{M} {}^{(K)}\sigma^2_{i,t-1|t-2} \, p_{i,t-1|t,t-1}$$

(6)

where $p_{i,t-1|t,t-1}$ is the augmented filtered probability as defined above. Consequently, here $^{(K)}\sigma^2_{t-1|t-2}$ becomes a function of $\Psi_{t-2}$ and $s_{t-1} = i$.

## 3  Filtering and Duality

In order to develop a theory of linear filtering for MS GARCH models, we need to link the model with some state space representations. Here we propose a state space representation and write the associated filter. Consider the model as in (1). For every $s_t = j$ and $s_{t-1} = i$, let us define $\varepsilon_t^2 = \sigma^2_{j,t} + v_t$, where $\sigma^2_{j,t} = \sigma^2_t(\Psi_{t-1}, s_t = j)$ and $v_t = \sigma^2_{j,t}(u_t^2 - 1)$. Then $(v_t)$ is zero mean serially uncorrelated. However, $(v_t)$ is not independent over time since it does not have a constant variance in time (i.e., it is not a homoskedastic process). Now we have

$$\varepsilon_t^2 = \sigma^2_{j,t} + v_t = \omega_j + \alpha_j \varepsilon^2_{t-1} + \beta_j \sigma^2_{i,t-1} + v_t = \omega_j + \alpha_j \varepsilon^2_{t-1} + \beta_j(\varepsilon^2_{t-1} - v_{t-1}) + v_t$$

where $\omega_j$, $\alpha_j$ and $\beta_j$ are the elements obtained by replacing $s_t$ by $j$ in $\omega_{s_t}$, $\alpha_{s_t}$ and $\beta_{s_t}$, respectively. So we can write a new representation of model (1) as

$$(1 - \delta_j L)\varepsilon_t^2 = \omega_j + (1 - \beta_j L)v_t \tag{7}$$

where $\delta_j = \alpha_j + \beta_j$ for $j = 1, \dots, M$. Now, setting $B_t = (\varepsilon_t^2 \quad v_t)'$, we get

$$\varepsilon_t^2 = \omega_j + (\delta_j \quad -\beta_j)\begin{pmatrix} \varepsilon^2_{t-1} \\ v_{t-1} \end{pmatrix} + v_t = \omega_j + (\delta_j \quad -\beta_j)B_{t-1} + v_t$$

for every $j = 1, \dots, M$. In order to simplify notations, let us define

$$y_t = \varepsilon_t^2, \quad H = (1 \quad 0), \quad F_{s_t} = \begin{pmatrix} \delta_{s_t} & -\beta_{s_t} \\ 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mu_{s_t} = \begin{pmatrix} \omega_{s_t} \\ 0 \end{pmatrix}.$$

Then, for every $s_t$, we obtain the following switching state space representation:

$$
\begin{cases}
y_t = HB_t \\
B_t = \mu_{s_t} + F_{s_t}B_{t-1} + Gv_t
\end{cases}
\tag{8}
$$

Representation (8) is similar but different to the switching dynamic model of [19] and [20]. The state vector $B_t$ is called the state (of the system) at time $t$; in line with [13], p.576, it can be partially unobservable as it is in this case. In fact, it includes the squared observed returns. Then, conditional on $s_{t-1} = i$ and $s_t = j$, we obtain the following filter:

*Prediction*

$$
B_{t|t-1}^{(i,j)} = \mu_j + F_j B_{t-1|t-1}^i
$$

$$
P_{t|t-1}^{(i,j)} = F_j P_{t-1|t-1}^i F_j' + GG'\sigma_{vj}^2
$$

$$
\eta_{t|t-1}^{(i,j)} = y_t - y_{t|t-1}^{(i,j)} = y_t - HB_{t-1|t-1}^{(i,j)}
$$

$$
f_{t|t-1}^{(i,j)} = HP_{t-1|t-1}^{(i,j)} H'
$$

*Updating*

$$
B_{t|t}^{(i,j)} = B_{t|t-1}^{(i,j)} + K_t^{(i,j)} \eta_{t|t-1}^{(i,j)}
$$

$$
P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} - K_t^{(i,j)} HP_{t|t-1}^{(i,j)}
$$

where $\sigma_{vj}^2 = var(v_t|\Psi_{t-1}, s_t = j)$ and $K_t^{(i,j)} = P_{t|t-1}^{(i,j)} H'[f_{t|t-1}^{(i,j)}]^{-1}$ is the Kalman gain

*Initial Conditions*

$$
B_{0|0}^j = (I_2 - F_j)^{-1}\mu_j = \begin{pmatrix} (1-\delta_j)^{-1}\omega_j \\ 0 \end{pmatrix}
$$

$$
\text{vec}(P_{0|0}^j) = \sigma_{vj}^2(I_4 - F_j \otimes F_j)^{-1}\text{vec}(GG') = \sigma_{vj}^2 \begin{pmatrix} (1-\delta_j^2)^{-1}(1-2\delta_j\beta_j+\beta_j^2) \\ 1 \\ 1 \\ 1 \end{pmatrix}
$$

$$
p(s_0 = i) = \pi_i \text{(steady-state probability).}
$$

Explicit derivation of the above filtering procedure is detailed in Appendix A2. Here $Y_{t-1} = \{y_{t-1}, \ldots, y_1\}$ is the information set up to time $t-1$, $B_{t|t-1}^i = E(B_t|Y_{t-1}, s_{t-1} = i)$ is an inference on $B_t$ based on $Y_{t-1}$ given $s_{t-1} = i$; $B_{t|t-1}^{(i,j)} = E(B_t|Y_{t-1}, s_t = j, s_{t-1} = i)$ is an inference on $B_t$ based on $Y_{t-1}$, given $s_t = j$ and $s_{t-1} = i$; $P_{t-1|t-1}^i$ is the mean squared error matrix of $B_{t-1|t-1}^i$ conditional on $s_{t-1} = i$; $P_{t|t-1}^{(i,j)}$ is the mean squared error matrix of $B_{t|t-1}^{(i,j)}$ conditional on $s_t = j$ and $s_{t-1} = i$; $\eta_{t|t-1}^{(i,j)}$ is the conditional forecast error of $y_t$ based on information up to time $t-1$, given $s_t = j$ and $s_{t-1} = i$; and $f_{t|t-1}^{(i,j)}$ is the conditional variance of forecast error $\eta_{t|t-1}^{(i,j)}$. Each iteration of the Kalman Filter produces an *M*-fold increase in the number of cases to consider. It is necessary to introduce some approximations to make the filter operable.

The key is to collapse the $(M \times M)$ posteriors $B_{t|t}^{(i,j)}$ and $P_{t|t}^{(i,j)}$ into $M$ posteriors $B_{t|t}^{j}$ and $P_{t|t}^{j}$. Hence, we mimic the approximation proposed by [20] and [19] applied to this state space representation (see Appendix A3). Let $B_{t|t}^{j}$ be the expectation based not only on $Y_t$ but also conditional on the random variable $s_t$ taking on the value $j$. Then the approximation results to be

$$B_{t|t}^{j} = \sum_{i=1}^{M} p_{i,t-1|t,t} \, B_{t|t}^{(i,j)}. \tag{9}$$

To justify the use of this approximation, note that the Kalman filter would give the conditional expectation if, conditional on $\Psi_{t-1}$ and on $s_t = j$, $s_{t-1} = i$, the distribution of $B_t$ is normal. However, the distribution of $B_t$, conditional on $\Psi_{t-1}$, $s_t = j$ and $s_{t-1} = i$, is a mixture of normal for $t > 2$. Hence, Kim proposes an approximation in which the exponential Gaussian mixture is collapsed down to $M$ Gaussians at each step. This is a natural proxy for such a process. Having such a convenient switching state space form associated to the initial MS GARCH, gives us the possibility to reconcile in a unique framework the estimation through linear filter or via auxiliary models. Duality exists when we modify the approximation described in (9) with different conditioning sets. The existence of this relationship between the approximated filter and collapsing procedures gives a theoretical ground in using the latter approaches. In fact, these approximated models of the MS GARCH were historically given based only on intuitive arguments. Now, from the measurement equation in (8) and using (9), we have

$$y_{t|t-1}^{j} = E(y_t|s_t = j, Y_{t-1}) = HE(B_t|s_t = j, Y_{t-1})$$

$$= H \sum_{i=1}^{M} p_{i,t-1|t,t-1} \, B_{t|t-1}^{(i,j)} = \sum_{i=1}^{M} p_{i,t-1|t,t-1} \, y_{t|t-1}^{(i,j)}$$

and

$$y_{t-1|t-1}^{j} = E(y_{t-1}|s_t = j, Y_{t-1}) = E(\sigma_{t-1}^2|Y_{t-1}, s_t = j)$$

$$= \sigma_{j,t-1|t-1}^2 = \sum_{i=1}^{M} p_{i,t-1|t,t-1} \, \sigma_{ij,t-1|t-2}^2.$$

In particular, if the conditional variance is not a function of $s_t = j$, we get

$$y_{t-1|t-1} = E(\varepsilon_{t-1}^2|Y_{t-1}) = E(\sigma_{t-1}^2|Y_{t-1})$$

$$= \sigma_{t-1|t-1}^2 = \sum_{i=1}^{M} p_{i,t-1|t,t-1} \, \sigma_{i,t-1|t-2}^2 \tag{10}$$

which coincides with ${}^{(K)}\sigma_{t-1}^2$ in Formula (6). Here ${}^{(K)}\sigma_{t-1}^2$ is only a function of $s_{t-1} = i$. Thus the approximation of the filter is dual to the one used as auxiliary model in [18]. This also means that if we change the conditioning scheme in (10), we obtain other auxiliary models. In fact, if we assume probabilities to be only

function of $s_{t-1} = i$ and if still $\sigma_{t-1}^2$ is a function of $s_{t-1}$ , we have the simplified Klaassen model. This gives the expression in (5), in fact:

$$^{(SK)}\sigma_{t-1}^2 = \sum_{i=1}^{M} {}^{(SK)}\sigma_{i,t-1|t-2}^2 \, p_{i,t-1|t-1} \cdot$$

Moreover, if we assume instead that $\sigma_{t-1}^2$ is a function of $s_{t-2} = k$ and also considering prediction probabilities of $s_{t-2} = k$, we get the auxiliary model proposed by [10]:

$$^{(D)}\sigma_{t-1}^2 = \sum_{i=1}^{M} {}^{(D)}\sigma_{k,t-1|t-2}^2 \, p_{k,t-2|t-1}$$

which is Equation (4). Finally, if we consider the conditioning set up to $Y_{t-2}$ rather than $Y_{t-1}$, we obtain

$$^{(G)}\sigma_{t-1}^2 = \sum_{i=1}^{M} {}^{(G)}\sigma_{i,t-1|t-2}^2 \, p_{i,t-1|t-2}$$

which is Equation (3) and corresponds to Gray's model. Hence, if we slightly change the conditioning set, we can obtain different specifications of the auxiliary models, moving from the state space form in (8). The usefulness of linking the filter to the four approximations is not only theoretical but also practical, providing a relatively easy method to conduct estimation. In particular, we have showed a direct connection between the filter and Klaassen collapsing procedure. However, an exercise to empirically prove the accuracy of the approximations can be found in [23], where it has been empirically estabilished that Klaassen model is the most effective one in generating consistent estimates for the path-dependent MS GARCH models.

## 4 An application on US Treasury Bill rates

We consider one-month US Treasury bill rates obtained from FRED for the period January 1970 trought April 1994 as in [14]. Figure 1 plots the data in level and in first difference. It is immediate the dramatic increase in interest rates that occurred during the Fed experiment and the OPEC oil crisis, which leads us to consider a 2-regime model. Then we fit a 2-state MS GARCH model as in (1) with both changes in regimes in the intercept term and in the persistence parameters of the volatility process. The values of the estimation are reported in Table 1 where estimated values along with robust standard errors are reported. In particular, the model estimated by filtering mimic with Kim and Nelson's (KN) approximation is labelled with Approximation 1 and it is dual to Klaassen (K). Note that Approximation 2 reproduces the auxiliary model of Gray (G) and values are in fact very close (see [14], Table 3, p.44). Finally, Approximation 3 and 4 are respectively dual to Dueker (D) and the simplified Klaassen (SK) models. The high-volatility regime

is characterized by more sentivity to recent shocks ($\hat{\alpha}_2 > \hat{\alpha}_1$) and less persistence ($\hat{\beta}_2 < \hat{\beta}_1$) than the low-volatility regime. Within each regime, the GARCH processes are stationary ($\hat{\alpha}_i + \hat{\beta}_i < 1$) and the parameter estimates suggest that the regimes are very persistent, so the source of volatility persistence will be important. In general, the four approximations are not very dissimilar to the others and the filtering algorithm has a very limited computational cost. Figure 2 plots smoothed probabilities $Pr(s_t = 1|\Psi_T)$ which are of interest to determine if and when the regime switching occurs. The plot of smoothed probabilities manages to identify crises periods that affected market indices. In particular, we could recognize three periods of high-variance. The first (1973-1975) corresponds to the OPEC oil crisis. The second is shorter and corresponds to the Fed experiment (1979-1983). The third is a short period around 1987, after the stock market crash.
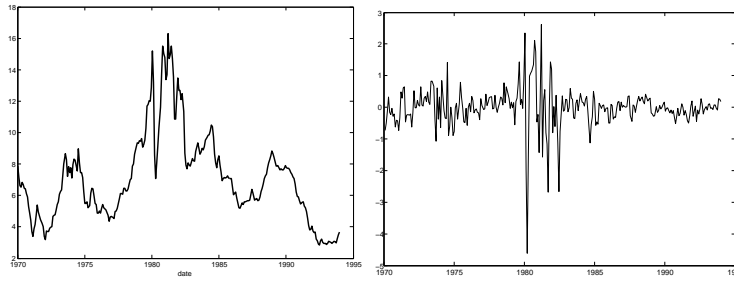


**Fig. 1** The left panel contains a time series plot of one-month US Treasury bill rates (in annualized percentage term). The sample period is from January 1970 to April 1994; a total of 1,267 observations. First differences of the series are shown in the right panel. The data are obtained from FRED database.

| **Approximations** | $\hat{p}$ | $\hat{q}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\omega}_1$ | $\hat{\omega}_2$ |
|---|---|---|---|---|---|---|---|---|
| Approximation 1 | 0.8467 | 0.9982 | 0.4113 | 0.0085 | 0.0184 | 0.4967 | 0.023 | 0.068 |
| (KN=K) | (0.1402) | (0.0334) | (0.0407) | (0.0927) | (0.0338) | (0.0331) | (0.2626) | (0.0908) |
| Approximation 2 | 0.8018 | 0.9157 | 0.391 | 0.0062 | 0.0203 | 0.4801 | 0.045 | 0.0713 |
| (=G) | (0.1146) | (0.0157) | (0.1624) | (0.2609) | (0.0640) | (0.1089) | (0.3591) | (0.2853) |
| Approximation 3 | 0.8467 | 0.9983 | 0.4112 | 0.0086 | 0.0184 | 0.4967 | 0.023 | 0.068 |
| (=D) | (0.1406) | (0.0336) | (0.0420) | (0.0923) | (0.0342) | (0.0337) | (0.2651) | (0.0799) |
| Approximation 4 | 0.8119 | 0.9160 | 0.397 | 0.0064 | 0.0212 | 0.4831 | 0.039 | 0.0692 |
| (=SK) | (0.1140) | (0.0157) | (0.1629) | (0.2618) | (0.0642) | (0.1092) | (0.3603 ) | (0.2867) |

**Table 1** Estimation of the parameters for 2-state MS GARCH model as in (1). Robust standard errors in parenthesis. $\hat{p}$ and $\hat{q}$ are the elements on the main diagonal of the transition probability matrix. The observables are one-month US Treasury bill rates (in annualized percentage term). The sample period is from January 1970 to April 1994; a total of 1,267 observations. The data are obtained from FRED database.
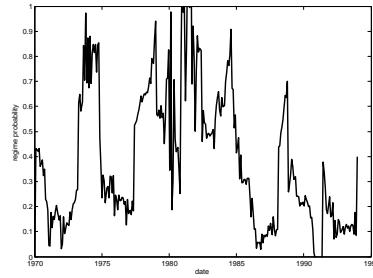
**Fig. 2** The panel shows MS GARCH smoothed probabilities of being in the high-volatility regime. Parameter estimates are based on a dataset of one-month Treasury Bill rates, reported in annualized percentage terms. The sample period is from January 1970 to April 1994; a total of 1,267 observations. The data are obtained from FRED database.

## 5 Conclusions

We deal with estimation of Markov Switching GARCH models. It is well-known that these models suffer of path-dependence, i.e., dependence of the entire path history of the data which makes Maximum Likelihood procedures unfeasible to apply. Therefore, we introduce filtering procedures based on approximated algorithms for switching state space representations. We show duality results in the estimation by approximated filtering method and auxiliary models proposed in the literature. Our filtering algorithm is readily programmable and with a very limited computational cost. An application on financial data shows the feasibility of the proposed approach.

## References

1. Augustyniak, M. (2014) "Maximum likelihood estimation of the Markov-switching GARCH model", *Computational Statistics & Data Analysis* **76**, 61-75.
2. Augustyniak, M., Boudreault, M., and Morales, M. (2017) "Maximum Likelihood Estimation of the Markov-Switching GARCH Model Based on a General Collapsing Procedure", in Methodology and Computing in Applied Probability, 1-24.
3. Bauwens, L., Preminger, A., and Rombouts, J. (2010) "Theory and Inference for a Markov Switching GARCH Model", *Econometrics Journal* **13**(2), 218-244.
4. Bauwens, L., Dufays, A., and Rombouts, J. V. (2014) "Marginal likelihood for Markov-switching and change-point GARCH models", *Journal of Econometrics* **178**, 508-522.
5. Billio, M., Casarin, R., and Osuntuy, A. (2014) "Efficient Gibbs Sampling for Markov Switching GARCH Models", forthcoming in *Computational Statistics and Data Analysis*.
6. Billio, M., Monfort A., and Robert, C.P. (1998a) "A MCMC approach to maximum likelihood estimation", *Prague Stochastics '98*, Vol.1, Ed. M.Huskova, P.Lachout and J.A.Visek, Union of Czach Mathematicians and Phusicists, 49-54.
7. Billio, M., Monfort A., and Robert, C.P. (1998b) "The simulated likelihood ratio (SLR) method", Document de Travail du CREST 9828, Paris.
8. Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* **52**, 5-59.

9. Cai, J. (1994) "A Markov model of unconditional variance in ARCH", *Journal of Business and Economic Statistics* **12**, 309-316.

10. Dueker, M.J. (1997) "Markov Switching in GARCH Processes and Mean Reverting Stock Market Volatility", *Journal of Business and Economic Statistics* **15**(1), 26-34.

11. Francq, C., Roussignol, M., and Zakoïan, J.M. (2001) "Conditional heteroskedasticity driven by hidden Markov chains", *Journal of Time Series Analysis* **22**(2), 197-220.

12. Francq, C., and Zakoïan, J.M. (2005) "The $L^2$-structures of standard and switching-regime GARCH models", *Stochastic Processes and their Applications* **115**, 1557-1582.

13. Gourieroux, C., and Monfort, A. (1997) "Time Series and Dynamic Models", *Cambridge University Press*, 1997.

14. Gray, S.F. (1996) "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process", *Journal of Financial Economics* **42**, 27-62.

15. Haas, M., Mittnik, S., and Paolella, M.S. (2004) "A new approach to Markov-Switching GARCH Models", *Journal of Financial Econometrics* **2**, 493-530.

16. Hamilton, J.D., and Susmel, R. (1994) "Autoregressive conditional heteroskedasticity and changes in regime", *Journal of Econometrics* **64**, 307-333.

17. Henneke, J.S., Rachev, S.T., Fabozzi, F.J., and Metodi, N. (2011) "MCMC-based estimation of Markov Switching ARMA–GARCH models", *Applied Economics* **43**(3), 259-271.

18. Klaassen, F. (2002) "Improving GARCH Volatility Forecasts with Regime-Switching GARCH", *Empirical Economics* **27**(2), 363-394.

19. Kim, C.J. (1994) "Dynamic Linear Models with Markov Switching", *Journal of Econometrics* **64**, 1-22.

20. Kim, C.J., and Nelson, C.R. (1999) "State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications", *The MIT Press*, 1999.

21. Lamoureux, C.G., and Lastrapes, W.D. (1990) "Persistence in variance, structural change, and the GARCH model", *Journal of Business and Economic Statistics* **8**, 225-234.

22. Nelson, D.B. (1990) "Stationarity and persistence in the GARCH(1,1) model", *Econometric Theory* **6**, 318-344.

23. Osuntuy, A.A. (2014), "Essays on Bayesian Inference with Financial Applications", PhD thesis, University Ca' Foscari of Venice.

# Appendix

**A1.** We show that $p_{t|t-1,t} = p(s_t|s_{t-1}, \Psi_t)$ can be espressed in terms of $p_{t|t-1,t-1}$ and the conditional density of $\varepsilon_t$ which depends on the current regime $s_t$ and the past regimes, i.e, $f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})$. In fact,

$$
\begin{aligned}
p_{t|t-1,t} = p(s_t|s_{t-1},\Psi_t) &= p(s_t|s_1,\ldots,s_{t-1},\Psi_t) = p(s_t|s_1,\ldots,s_{t-1},\varepsilon_t,\Psi_{t-1}) \\
&= \frac{f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})p(s_t|s_1,\ldots,s_{t-1},\Psi_{t-1})}{f(\varepsilon_t|s_1,\ldots,s_{t-1},\Psi_{t-1})} \\
&= \frac{f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})p(s_t|s_{t-1},\Psi_{t-1})}{f(\varepsilon_t|s_1,\ldots,s_{t-1},\Psi_{t-1})} = \frac{f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})p_{t|t-1,t-1}}{f(\varepsilon_t|s_1,\ldots,s_{t-1},\Psi_{t-1})}
\end{aligned}
$$

where $f(\varepsilon_t|s_1,\ldots,s_{t-1},\Psi_{t-1}) = \sum_{s_t=1}^{M} f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})p(s_t|s_{t-1},\Psi_{t-1})$ $= \sum_{s_t=1}^{M} f(\varepsilon_t|s_1,\ldots,s_t,\Psi_{t-1})\, p_{t|t-1,t-1}$.

**A2.** We present explicit derivation of the filter for MS GARCH as given in Section 3. The prediction step is obtained as follows

$$
\begin{aligned}
B_{t|t-1}^{(i,j)} = E(B_t|\Psi_{t-1},s_t=j,s_{t-1}=i) &= E(\mu_{s_t}+F_{s_t}B_{t-1}+Gv_t|\Psi_{t-1},s_t=j,s_{t-1}=i) \\
&= \mu_j + F_j E(B_{t-1}|\Psi_{t-1},s_t=j,s_{t-1}=i) = \mu_j+F_j B_{t-1|t-1}^i.
\end{aligned}
$$

In particular, we have $B_t - B_{t|t-1}^{(i,j)}|_{s_t=j} = F_j(B_{t-1} - B_{t-1|t-1}^i) + Gv_t$. Then

$$
\begin{aligned}
P_{t|t-1}^{(i,j)} &= E[(B_t - B_{t|t-1}^{(i,j)})(B_t - B_{t|t-1}^{(i,j)})'|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\
&= E[(F_j(B_{t-1} - B_{t-1|t-1}^i) + Gv_t)(F_j(B_{t-1} - B_{t-1|t-1}^i) + Gv_t)'|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\
&= F_j E[(B_{t-1} - B_{t-1|t-1}^i)(B_{t-1} - B_{t-1|t-1}^i)'|\Psi_{t-1}, s_{t-1} = i]F_j' + GE(v_t^2|s_t = j)G' \\
&= F_j P_{t-1|t-1}^i F_j' + GG'\sigma_{vj}^2
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_{t|t-1}^{(i,j)} &= y_t - y_{t|t-1}^{(i,j)} = y_t - E(y_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) = y_t - E(HB_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) \\
&= y_t - HE(B_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) = y_t - HB_{t|t-1}^{(i,j)}.
\end{aligned}
$$

Hence, $\eta_{t|t-1}^{(i,j)}|_{\Psi_{t-1}, s_t=j, s_{t-1}=i} = H(B_t - B_{t|t-1}^{(i,j)}) + v_t$ and

$$
\begin{aligned}
f_{t|t-1}^{(i,j)} &= E[(\eta_{t|t-1}^{(i,j)})^2|\Psi_{t-1}, s_t = j, s_{t-1} = i)] \\
&= E[(H(B_t - B_{t|t-1}^{(i,j)}))(H(B_t - B_{t|t-1}^{(i,j)}))'|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\
&= HE[(B_t - B_{t|t-1}^{(i,j)})(B_t - B_{t|t-1}^{(i,j)})'|\Psi_{t-1}, s_t = j, s_{t-1} = i]H' = HP_{t|t-1}^{(i,j)}H'.
\end{aligned}
$$

Furthermore, the updating step is derived as follows. Define $Z_1 = B_t$ and $Z_2 = \eta_{t|t-1}^{(i,j)} = y_t - y_{t|t-1}^{(i,j)}$. Then $\mu_1 = E[Z_1|\Psi_{t-1}, s_t = j, s_{t-1} = i] = B_{t|t-1}^{(i,j)}$, $\mu_2 = E[Z_2|\Psi_{t-1}, s_t = j, s_{t-1} = i] = 0$, $\Sigma_{11} = P_{t|t-1}^{(i,j)}$ and $\Sigma_{22} = f_{t|t-1}^{(i,j)}$. We have

$$
\begin{aligned}
\Sigma_{12} = cov(Z_1, Z_2) &= E[(B_t - B_{t|t-1}^{(i,j)})\eta_{t|t-1}^{(i,j)}|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\
&= E[(B_t - B_{t|t-1}^{(i,j)})(B_t - B_{t|t-1}^{(i,j)})'H'|\Psi_{t-1}, s_t = j, s_{t-1} = i] = P_{t|t-1}^{(i,j)}H' = \Sigma_{21}'.
\end{aligned}
$$

Thus $Z_1|_{Z_2,\Psi_{t-1}, s_t=j, s_{t-1}=i}$ is given by $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2)$, that is, $B_{t|t}^{(i,j)} = B_{t|t-1}^{(i,j)} + P_{t|t-1}^{(i,j)}H'[f_{t|t-1}^{(i,j)}]^{-1}\eta_{t|t-1}^{(i,j)}$. Further, we have $\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, hence $P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} - K_t^{(i,j)}HP_{t|t-1}^{(i,j)}$, where $K_t^{(i,j)} = P_{t|t-1}^{(i,j)}H'[f_{t|t-1}^{(i,j)}]^{-1}$ is the Kalman gain.

**A3.** Here we derive the approximation on the line of [20] applied to model in (8), which is Equation (9):

$$
\begin{aligned}
B_{t|t}^j &= \frac{\sum_{i=1}^M B_{t|t}^{(i,j)} p(s_{t-1} = i, s_t = j|Y_t)}{p(s_t = j|Y_t)} = \sum_{i=1}^M \frac{p(s_{t-1} = i, s_t = j|Y_t)}{p(s_t = j|Y_t)} B_{t|t}^{(i,j)} \\
&= \sum_{i=1}^M p(s_{t-1} = i|s_t = j, Y_t) B_{t|t}^{(i,j)} = \sum_{i=1}^M p_{i,t-1|t,t} B_{t|t}^{(i,j)}.
\end{aligned}
$$