Ca' Foscari University of Venice

Department of Economics

Working Paper

Monica Billio and Maddalena Cavicchioli

Markov switching models for volatility: Filtering, approximation and duality

ISSN: 1827-3580
No. 24/WP/2013

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 24/WP/2013
ISSN 1827-3580

MARKOV SWITCHING MODELS FOR VOLATILITY:

FILTERING, APPROXIMATION AND DUALITY

Monica Billio and Maddalena Cavicchioli

Cà Foscari University of Venice, Italy

30/10/2013

Abstract. This paper is devoted to show duality in the estimation of Markov Switching (MS) processes for volatility. 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 MS-GARCH model in a suitable linear State Space representation, we are able to give a unique framework to reconcile the estimation obtained by the Kalman Filter and with some auxiliary models proposed in the literature. Reasoning in the same way, we present a linear Filter for MS-Stochastic Volatility (MS-SV) models on which different conditioning sets yield more flexibility in the estimation. Estimation on simulated data and on short-term interest rates shows the feasibility of the proposed approach.

Keywords. Markov Switching, MS-GARCH model, MS-SV model, estimation, auxiliary model, Kalman Filter.

JEL Codes. C01, C13, C58.

Address for correspondence:
Maddalena Cavicchioli
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: maddalena.cavicchioli@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/nqcontent.cfm?a_id=86302)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

1. Introduction

Time varying volatility is one of the main property of economic time series, common especially to many financial time series. Moreover, describing and, where possible, forecasting volatility is a key aspect in financial economics and econometrics. It is not only a statistical exercise but it has also important impacts in terms of asset allocation, asset pricing as well as value-at-risk computation and thus for risk management. A lot of work has been done on two popular classes of models which describe time-varying volatility: Generalized Autoregressive Conditional Heteroschedasticity (GARCH)-type models and Stochastic Volatility (SV)-type models. GARCH models (Bollerslev (1986), Nelson (1990), Lamoureux and Lastrapes (1990)) are commonly known as observation-driven models (see Shephard (1996)). In fact, they describe the variance as a linear function of the squares of past observations and then one type of shock alone drives both the series itself and its volatility. On the contrary, SV models (Taylor (1986), Harvey, Ruiz and Shephard (1994)) belong to the class of parameter-driven models since these models are driven by two type of shocks, one of which influences the volatility. The presence of unobserved or latent components makes SV models harder to estimate and to handle statistically, while GARCH parameters can easily be estimated using maximum likelihood procedure. In the latter models, one potential source of misspecification is that the structural form of conditional means and variances is relatively inflexible and it is held fixed throughout the sample period. In this sense, they are called *single-regime* models since a single structure is assumed for the conditional mean and variance.

In order 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 that generates the financial serie undergoes a finite number of changes over the sample period. These coefficients are unknown and must be estimated, and, althought 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 GARCH models is the path dependence. Cai (1994) and Hamilton and Susmel (1994) have argued that MS-GARCH models are essentially intractable and impossible to estimate due to the dependence of conditional variance 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 of 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.

In the first part of this paper, we will consider the univariate version of MS-GARCH and some methods proposed to bypass the problem of path dependence. 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 allow similar effects of the original one. Models which elude in this way the path dependence problem are proposed by Gray (1996), Dueker (1997) and Klaassen (2002), among others. Gray (1996) 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 Dueker (1997), with a slightly different approach. He extends the information set including also current information on the considered series. Furthermore, Klaassen (2002) puts further this idea. Particularly, when integrating out the unobserved regimes, he uses all available information, whereas Gray uses only part of it. Another method to deal with MS-GARCH models has been proposed by Haas, Mittnik and Paolella (2004) for which the variance is disaggregated in independent processes; this is a simple generalization of the GARCH process to a multi-regime setting. Finally, Bayesian approach based on Markov Chain Monte Carlo (MCMC) Gibbs technique for estimating MS-GARCH can be found in Bauwens, Preminger and Rombouts (2010) and Bauwens, Dufays and Rombouts (2011), Henneke, Rachev, Fabozzi and Metodi (2011) or Billio, Casarin and Osuntuyi (2012). Other approaches based on both Monte Carlo methods combined with expectation-maximization algorithm and importance sampling to evaluate ML estimators can be found in Augustyniak (2013) and Billio, Monfort and Robert (1998a and 1998b).

In the second part of the paper, we will consider the extension of univariate SV model with regime-switching features. If SV models are difficult to estimate due to the latent variable, MS-SV are even more complicated because there are two hidden levels in the latent structure. So, MS-SV models have been studied and estimated mainly with Bayesian techniques. For example, So, Lam and Li (1998) adopt MCMC method and they construct Bayesian estimators by Gibbs sampling. Another Bayesian approach is sequential simulation based filtering (Particle Filter). See, for instance, Casarin (2004) and Carvalho and Lopes (2007).

The main contribution of the present paper is to give a unique framework to reconcile the estimation obtained by the above auxiliary models from one side, and Kim's (1994) filtering algorithm for Markov switching state space from the other. Kim's algorithm can be used, under some regularity conditions, to obtain inferences about any dynamic time series model with Markov switching that can be put in a state space form. It is a very flexible approach and allows the estimation of a broad class of models. However, to make the filter operable, at each iteration it collapses M^2 posteriors (where M is the number of states) in M of it, employing an approximation. Finally, Quasi Maximum Likelihood estimation of the model recovers the unknown parameters. Then our first contribution is to show duality in the estimation of Markov Switching processes for volatility. In particular, having a suitable linear state space representation for the MS-GARCH model, we are able to prove the equivalence in the estimation obtained by Kim's Filter and through auxiliary models proposed in the literature. The second contribution relates instead to MS-SV models. In fact, we are able to extend the approach previously used for MS-GARCH to MS-SV models. In particular, we parallel the model with the gaussian state space model and we propose a linear Filter on which different conditioning information sets yield more flexibility in the estimation. Numerical and empirical applications show the feasibility of these approaches.

The paper is structured as follows. In Section 2 we specify the MS-GARCH model of interest and introduce some concepts and notations. Section 3 reviews the main auxiliary models for MS-GARCH which are proposed in the literature to overpass the path dependence problem. In Section 4 we present a linear state space representation associated to the MS-GARCH and determine the algorithm for the linear filter. This serves to prove our duality results discussed in Section 5. In Section 6 we write a linear approximated filter for MS-SV models. In Section 7 we compare estimation of the parameters using different approximations in the proposed filter for simulated data and short-term interest rates. Section 8 concludes. Finally, Appendix A describes in details some Formulae and in Appendix B we recall the main results about the stationarity of Markov Switching models and particulary applied to our specifications.

2. Markov Switching GARCH

Let ϵ_t be the observed univariate[1] time series variable (as for instance, returns on a financial asset) centered on its mean and let s_t be a discrete, unobserved state variable with M-states. The Markov Switching GARCH(1,1) model is defined as

$$(1) \qquad \begin{cases} \epsilon_t = \sigma_t(\Psi_{t-1}, \theta(s_t))u_t \\ \sigma_t^2(\Psi_{t-1}, s_t) = \omega(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2(\Psi_{t-2}, s_{t-1}) \end{cases}$$

[1]The proposed setting can be easily extended to a multivariate framework. This can be done on the line of multivariate GARCH models to regime-switching framework proposed by Billio and Caporin (2005) and Pellittier (2006). However, note that multivariate volatility models in the context of single regime switching are the Constant Conditional Correlation (CCC) model of Bollerslev (1990) and the Dynamic Conditional Correlation (DCC) model of Engle (2002).

where $u_t \sim IID(0,1)$, $\omega(s_t) > 0$, $\alpha(s_t), \beta(s_t) \geq 0$ and $\theta(s_t)$ is the parameter vector defined as $\theta(s_t) = (\omega(s_t), \alpha(s_t), \beta(s_t))^{'}$. Here $\Psi_{t-1} = \{\epsilon_{t-1}, \ldots, \epsilon_1\}$ denotes the information set of observations available up to time $t-1$. Moreover, s_t is a M-state first order Markov chain with *transition probabilities*, which are assumed time invariant[2],

$$\pi_{ij,t} = p(s_t = j | s_{t-1} = i)$$

where

$$\sum_{j=1}^{M} \pi_{ij,t} = 1$$

for every $i = 1, \ldots, M$.
Let us introduce the following concepts and notations:

- $p(s_t = j | \Psi_{t-1}) = p_{j,t|t-1}$ which is the *prediction probability*;
- $p(s_t = j | \Psi_t) = p_{j,t|t}$ which is the *filtered probability*.

From these we can compute the *augmented filtered probability* as

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

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 ϵ_t which depends on the current regime s_t and all past regimes, i.e, $f(\epsilon_t | s_1, \ldots, s_t, \Psi_{t-1})$. Computation details are shown in Appendix A1.

3. Auxiliary Models for MS-GARCH

As argued in the Introduction, the main problem to face when dealing with the estimation of Markov Switching GARCH model is the path dependence, which is the dependence of the conditional variance on the entire sequence of regimes. 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 the MS-GARCH process

$$\sigma_t^2(\Psi_{t-1}, s_t) = \omega(s_t) + \alpha(s_t)^{(SP)} \epsilon_{t-1}^2 + \beta(s_t)^{(SP)} \sigma_{t-1}^2. \tag{2}$$

[2]If the information variables that govern time-variation in the transition probabilities is conditionally uncorrelated with the state of the Markov process, which holds in general, Hamilton's (1989) filtering method is still valid also with time-varying transition probabilities.

In the literature there are different specifications (in short, SP) of ${}^{(SP)}\epsilon^2_{t-1}$ and ${}^{(SP)}\sigma^2_{t-1}$ which in turn define different approximations of the original process. In this Section we give a detailed description of four auxiliary models presented in the literature, specifying the superscript in (2) with the initial letter of the author who proposed that specification.

3a. Gray's Model

The first attempt to eliminate the path dependence is proposed by Gray (1996). He approximates the original model by replacing the lagged conditional variance σ^2_{t-1} with a proxy ${}^{(G)}\sigma^2_{t-1}$ as follows:

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

where, according to the model, ${}^{(G)}\sigma^2_{t-1|t-2}$ turns out to be a function of Ψ_{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.

3b. Dueker's Model

In the previous approximation, the information coming from ϵ_{t-1} is not used. Dueker (1997) proposes to change the conditioning scheme including ϵ_{t-1} while assuming that σ^2_{t-1} is a function of Ψ_{t-2} and s_{t-2}. Hence

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

so that ${}^{(D)}\sigma^2_{t-1|t-2}$ is a function of Ψ_{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

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

3c. Simplified Klaassen's Model

The approximation proposed by Klaassen (2002) is similar to that from Dueker (1997) but it assumes that σ^2_{t-1} is a function of Ψ_{t-2} and s_{t-1}. So it results computationally simpler. In fact, we have

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

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

3d. Klaassen's Model

Finally, Klaassen (2002) generalizes the previous auxiliary model including in the conditioning set the information also coming from the current regime s_t. So σ^2_{t-1} turns out to be approximated as

$$
\begin{aligned}
{}^{(K)}\sigma^2_{t-1} &= E[\sigma^2_{t-1}(\Psi_{t-2}, s_{t-1})|\Psi_{t-1}, s_t = j] \\
&= \sum_{i=1}^{M} \sigma^2_{t-1}(\Psi_{t-2}, s_{t-1}=i)\; p(s_{t-1}=i|\Psi_{t-1}, s_t=j) \qquad (6)\\
&= \sum_{i=1}^{M} {}^{(K)}\sigma^2_{i,t-1|t-2}\; p_{ij,t-1|t,t-1}
\end{aligned}
$$

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

4. State Space Representation and Filtering

In order to develop a theory of linear filtering for MS-GARCH models, we need to associate to the model some linear state space representations. In this Section we propose a state space representation and write the associated Kalman Filter. For this purpose, we use notations from Kim (1994) and Kim and Nelson (1999) which study Markov switching state space models. They propose basic filtering and smoothing algorithms, along with maximum likelihood estimation, for a broad class of Markov switching models which can be written in state space form. This linear filter can be used, under some regularity conditions, to obtain approximate inferences. In fact, it introduces an approximation by collapsing information on the regimes story at each iteration. Such an approximation will be presented hereafter.

Consider the model as in (1). For every $s_t = j$ and $s_{t-1} = i$, let us define $\epsilon_t^2 = \sigma_{j,t}^2 + v_t$, where $\sigma_{j,t}^2 = \sigma_t^2(\Psi_{t-1}, s_t = j)$ and $v_t = \sigma_{j,t}^2(u_t^2 - 1)$. Then v_t is a white noise with zero mean and variance σ_{vj}^2 and $v_t \in [-\sigma_{j,t}^2, +\infty[$. Now we have

$$\begin{aligned}\epsilon_t^2 &= \sigma_{j,t}^2 + v_t \\ &= \omega_j + \alpha_j \epsilon_{t-1}^2 + \beta_j \sigma_{i,t-1}^2 + v_t \\ &= \omega_j + \alpha_j \epsilon_{t-1}^2 + \beta_j(\epsilon_{t-1}^2 - v_{t-1}) + v_t.\end{aligned}$$

where ω_j, α_j and β_j are the elements obtained by replacing s_t by j in ω_{s_t}, α_{s_t} and β_{s_t}, respectively.
So we can write the MS-ARMA(1,1) representation of the process in (1) as

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

where $\delta_j = \alpha_j + \beta_j$ for $j = 1, \ldots, M$. See, for example, Gourieroux and Monfort (1997). For stationarity conditions concerning with such a process we refer to Appendix B.
Setting $B_t = \begin{pmatrix} \epsilon_{t-1}^2 \\ v_{t-1} \end{pmatrix}$, we get

$$\epsilon_t^2 = \omega_j + (\delta_j \quad -\beta_j) \begin{pmatrix} \epsilon_{t-1}^2 \\ v_{t-1} \end{pmatrix} + v_t = \omega_j + (\delta_j \quad -\beta_j) B_t + v_t$$

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

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

Then, for every s_t, we obtain the following state space representation [3]:

$$\begin{cases} y_t = \omega_{s_t} + H_{s_t} B_t + v_t \\ B_t = \mu_{s_t} + F_{s_t} B_{t-1} + G v_{t-1} \end{cases} \tag{8}$$

[3]Note that other state space representations can be associated to the model in (1). For instance, following the line of Kim and Nelson (1999), Example 2, Chapter 3. Our choice tends to be the less restrictive in term of stationarity conditions, hence more general.

Conditional on $s_{t-1} = i$ and $s_t = j$, the Kalman Filter is:
Prediction

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

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

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

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

Updating

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

where $K_t^{(i,j)} = P_{t|t-1}^{(i,j)} H_j^{'} [f_{t|t-1}^{(i,j)}]^{-1}$ is the Kalman gain

Initial Conditions

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

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

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

So $Y_{t-1} = \{y_{t-1}, \ldots, y_1\}$ is the information set up to time $t-1$, $B_{t-1|t-1}^{i} = E(B_t | Y_{t-1}, s_{t-1} = i)$ 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 consider the approximation proposed by Kim and Nelson (1999) and Kim (1994) applied to this state space representation (explicit computations are in Appendix A2) . Let $B^j_{t|t}$ be the expectation based not only on Y_t but also conditional on the random variable s_t taking on the value j. Then

$$B^j_{t|t} = \sum_{i=1}^{M} p_{ij,t-1|t,t} \; B^{(i,j)}_{t|t}. \tag{9}$$

5. Duality Results

Having such a convenient switching state space form associated to the initial MS-GARCH, gives us the possibility to reconcile in an unique framework the estimation through linear filter as described in Section 4 or via auxiliary models presented in Section 3. Duality exists when modifying the approximation described in (9) with different conditioning sets. From the measurement equation in (8) and using (9), we get

$$\begin{aligned} y^j_{t|t} &= E(y_t|s_t = j, Y_t) = \omega_j + H_j \; B^j_{t|t} \\ &= \omega_j + H_j \; \sum_{i=1}^{M} p_{ij,t-1|t,t} \; B^{(i,j)}_{t|t} \\ &= \sum_{i=1}^{M} p_{ij,t-1|t,t} \; (\omega_j + H_j \; B^{(i,j)}_{t|t}) \\ &= \sum_{i=1}^{M} p_{ij,t-1|t,t} \; y^{(i,j)}_{t|t} \end{aligned}$$

as $\sum_{i=1}^{M} p_{ij,t-1|t,t} = 1$. Here the expectation operator is meant in the sense of Kim and Nelson's book (1999). In the same way we can obtain

$$y^j_{t|t-1} = E(y_t|Y_{t-1}, s_t = j) = \sum_{i=1}^{M} p_{ij,t-1|t,t-1} \; y^{(i,j)}_{t|t-1}$$

and

$$\begin{aligned} y^j_{t-1|t-1} &= E(y_{t-1}|Y_{t-1}, s_t = j) = E(\sigma^2_{t-1}|Y_{t-1}, s_t = j) \\ &= \sigma^2_{j,t-1|t-1} = \sum_{i=1}^{M} p_{ij,t-1|t,t-1} \; \sigma^2_{ij,t-1|t-2}. \end{aligned}$$

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

$$\begin{aligned} y_{t-1|t-1} &= E(\epsilon^2_{t-1}|Y_{t-1}) = E(\sigma^2_{t-1}|Y_{t-1}) \\ &= \sigma^2_{t-1|t-1} = \sum_{i=1}^{M} p_{ij,t-1|t,t-1} \; \sigma^2_{i,t-1|t-2} \end{aligned} \tag{10}$$

which coincides with ${}^{(K)}\sigma^2_{t-1}$ in Formula (6). Here ${}^{(K)}\sigma^2_{t-1}$ is only a function of $s_{t-1} = i$. Thus the approximation of the Kalman Filter is dual to the one used as auxiliary model from Klaassen (2002). This also means that if we change the conditioning scheme in (10), we obtain others auxiliary models. In fact, if we assume probabilities to be only function of $s_{t-1} = i$ and if still σ^2_{t-1} is a function of s_{t-1} , we have the Simplified Klaassen's model (2002). This gives the expression in (5), in fact:

$$ {}^{(SK)}\sigma^2_{t-1} = \sum_{i=1}^{M} {}^{(SK)}\sigma^2_{i,t-1|t-2}\; p_{i,t-1|t-1}\,. $$

Moreover, if we assume instead that σ^2_{t-1} 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 Dueker (1997):

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

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

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

which is Formula (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). To conclude, this proves ambivalence in the estimation via Kalman Filter and via approximated models. In Section 7, we will show the feasibility of the filtering procedure through numerical and empirical applications.

6. Markov Switching Stochastic Volatility

When we consider Markov Switching Stochastic Volatility model and in general parameter-driven models, we are facing a double level of latency which makes estimation and statistical analysis harder. However, there are very good reason to investigate this kind of models, as for instance, easier properties or generalization to the multivariate case as well as continuous time counterpart. Then, we consider the following MS-SV model

$$ \begin{cases} \epsilon_t = \exp\{\frac{1}{2}h_t\}u_t \\ h_t = \mu_{s_t} + \rho_{s_t}h_{t-1} + v_t \end{cases} \tag{11} $$

where $u_t \sim IIN(0,1)$ and $v_t \sim IIN(0, \sigma^2_{v_{s_t}})$. Here the error terms are assumed to be independent of one other. To discuss stationarity conditions of the process, we will later rewrite the model in MS-ARMA form and stationarity conditions are discussed at the end of Appendix B.

Following Harvey, Ruiz and Shephard (1994), we easily obtain a linear state space form. In fact, squaring (11) and taking logs, we have

$$\log \epsilon_t^2 = \alpha + h_t + e_t$$

where $\alpha = E(\log u_t^2) \cong -1.270$ and $e_t = \log u_t^2 - E(\log u_t^2)$. Thus $e_t \sim IIN(0, \frac{\pi^2}{2})$, where $\frac{\pi^2}{2} \cong 4.935$, and higher moments of (e_t) are known. Now, replacing $\log \epsilon_t^2$ with y_t, the MS-SV can be written as

$$\begin{cases} y_t = \alpha + h_t + e_t \\ h_t = \mu_{s_t} + \rho_{s_t} h_{t-1} + v_t \end{cases} \tag{12}$$

Hence, it is natural to propose the Kalman filter for model (12) following the line of Kim and Nelson (1999, Chapter 5). In this case, conditional on $s_t = j$ and $s_{t-1} = i$, we get

Prediction

- $h_{t|t-1}^{(i,j)} = \mu_j + \rho_j h_{t-1|t-1}^i$
- $P_{t|t-1}^{(i,j)} = \rho_j^2 P_{t-1|t-1}^i + \sigma_{vj}^2$
- $\eta_{t|t-1}^{(i,j)} = y_t - y_{t|t-1}^{(i,j)} = y_t - h_{t|t-1}^{(i,j)} - \alpha$
- $f_{t|t-1}^{(i,j)} = P_{t|t-1}^{(i,j)} + \frac{\pi^2}{2}$

Updating

- $h_{t|t}^{(i,j)} = h_{t|t-1}^{(i,j)} + P_{t|t-1}^{(i,j)}[f_{t|t-1}^{(i,j)}]^{-1}\eta_{t|t-1}^{(i,j)}$
- $P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} - P_{t|t-1}^{(i,j)}[f_{t|t-1}^{(i,j)}]^{-1}P_{t|t-1}^{(i,j)}$

where $K_t^{(i,j)} = P_{t|t-1}^{(i,j)}[f_{t|t-1}^{(i,j)}]^{-1}$ is the Kalman gain.

Initial Conditions

- $h_{0|0}^j = \mu_j(1-\rho_j)^{-1}$
- $P_{0|0}^j = \sigma_{vj}^2(1-\rho_j^2)^{-1}$
- $p(s_0 = i) = \pi_i$ (steady-state probability).

If we apply the approximation proposed by Kim and Nelson (1999) to this state space representation, we can write

$$h^{j}_{t|t} = \sum_{i=1}^{M} h^{(i,j)}_{t|t} \; p_{ij,t-1|t,t}.$$

At this point, different conditioning sets can be applied to the above approximation, mimic the same ideas used to obtain different auxiliary models in the MS-GARCH model. In the sequel we propose four approximations for the MS-SV in (12). Approximation 1 denotes Kim and Nelson's approximation as specified above. As done for the MS-GARCH, if we change the conditioning set we can obtain different and possibily more precise estimates. Approximation 2 changes the conditioning set on the volatility up to $t-1$:

$$h^{j}_{t|t} = \sum_{i=1}^{M} h^{(i,j)}_{t|t-1} \; p_{ij,t-1|t,t}.$$

Approximation 3 consider the information set up to Y_{t-1} only for the augmented filtered probabilities:

$$h^{j}_{t|t} = \sum_{i=1}^{M} h^{(i,j)}_{t|t} \; p_{ij,t-1|t,t-1}.$$

The last approximation 4 simultaneously has the features of 2 and 3, conditioning both volatility and probabilities at $t-1$:

$$h^{j}_{t|t} = \sum_{i=1}^{M} h^{(i,j)}_{t|t-1} \; p_{ij,t-1|t,t-1}.$$

In the next Section we will test these specifications in a simulated study in order to investigate differences in the implementation of the Filter.

7. Applications

In this Section we apply the methods described above both to Monte Carlo experiment and real data. In particular, the aim of these applications is to show the feasibility of the proposed approaches via linear filtering for both Markov switching GARCH and SV models. Note that this method has the advantage of avoiding fine-tuning procedures implemented in most Bayesian estimation techniques. In fact, giving some initial conditions, the only duty of the researcher is to decide which approximation to adopt in the filtering procedure.

7a. Simulation study

In this Subsection, we draw some comparisons from a simulation study performed by So, Lam and Li (1998). In that paper, they simulate a Markov switching stochastic volatility model with three states and parameters described hereafter and estimate the model through MCMC procedure and Gibbs sampler. Thus the model is a MS(3)-SV as in (11) with fixed ρ equal to 0.5, $v_t \sim N(0, 0.2)$ and the intercept equal to

$$\mu_{s_t} = \begin{cases} -1 & \text{if} \quad s_t = 1 \\ -2 & \text{if} \quad s_t = 2 \\ -5 & \text{if} \quad s_t = 3 \end{cases} .$$

The state variables are generated by a first order Markov process with transition probability matrix

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.9 & 0 & 0.05 \\ 0 & 0.95 & 0.05 \\ 0.1 & 0.05 & 0.9 \end{pmatrix}$$

which implies high persistence in each regime. A dataset of $n = 400$ observations has been simulated from the model. We estimate the model with the filter proposed in Section 6 and 2,000 iterations are considered.
Results are summarized in Table 1 where means and standard deviations are given, together with the Bayesian estimators of So, Lam and Li and true values. As point estimate, all the approximated filters give close values to the corresponding true one. The persistence parameter ρ is better captured by Approximation 3 or 4, which is also the faster; those seem to be the best choices. Finally, our estimates obtained via Kalman filters give closer result to the true values with respect to the Bayesian counterpart. In fact, the Mean Square Error (MSE) value for the third approximation is equal to 0.00203 and the MSE computed by So, Lam and Li is 0.00330.

7b. Real Data: an application on US Treasury Bill rates

As a second application, we use real data and the same dataset as in Gray (1996). The data are one-month US Treasury bill rates obtained from FRED for the period January 1970 trought April 1994. Figure 1 plots the data. 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 regimes model.
Then we fit the model in (1) as MS(2)-GARCH and in (11) as MS(2)-SV 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 3 and 4, respectively. Table 3 describes

Estimation	$\hat{p}_{00}$	$\hat{p}_{10}$	$\hat{p}_{01}$	$\hat{p}_{11}$	$\hat{p}_{02}$	$\hat{p}_{12}$	$\hat{\rho}$	$\hat{\sigma}_v^2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	**Time***
Approximation 1	0.9489	0.0000	0.0000	0.9535	0.0518	0.0520	0.4838	0.2083	-0.8836	-2.0194	-4.9357	1.86
	(0.1165)	(0.0001)	(0.0001)	(0.0167)	(0.0009)	(0.0023)	(0.0207)	(0.0034)	(0.1073)	(0.0226)	(0.1909)	
Approximation 2	0.9263	0.0010	0.0000	0.9197	0.0540	0.0598	0.4858	0.2486	-0.8299	-2.096	-5.0332	1.92
	(0.0647)	(0.0005)	(0.0019)	(0.0689)	(0.0113)	(0.0217)	(0.1105)	(0.0697)	(0.3499)	(0.3628)	(0.5486)	
Approximation 3	0.9529	0.0011	0.0000	0.9334	0.0560	0.0471	0.5016	0.2279	-0.8659	-2.0545	-4.9824	1.83
	(0.0256)	(0.0017)	(0.0076)	(0.0912)	(0.0253)	(0.0212)	(0.0463)	(0.1027)	(0.5143)	(0.6148)	(0.7895)	
Approximation 4	0.9529	0.0011	0.0000	0.9334	0.0560	0.0471	0.5016	0.2279	-0.8659	-2.0545	-4.9824	1.83
	(0.0256)	(0.0017)	(0.0076)	(0.0912)	(0.0253)	(0.0212)	(0.0463)	(0.1027)	(0.5143)	(0.6148)	(0.7895)	
So, Lam and Li	0.877	0.026	0.027	0.941	0.079	0.040	0.513	0.314	-0.916	-2.071	-5.005	–
	(0.049)	(0.040)	(0.031)	(0.040)	(0.031)	(0.024)	(0.059)	(0.160)	(0.231)	(0.167)	(0.332)	
True values	0.95	0	0	0.95	0.05	0.05	0.5	0.2	-1	-2	-5	–

Table 1: Estimation results of simulated data from model (11). Robust standard errors in parenthesis.
* Minutes for single iteration. The sample period is from January 1970 to April 1994; a total of 1.267 observations. The data are obtained by the FRED database.

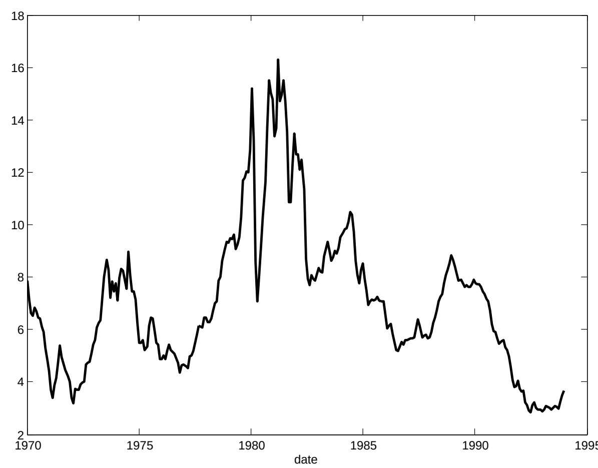

Figure 1: The panel contains a time series plot of monthly one-month US Treasury bill rates (in annualized percentage term). The sample period is from January 1970 to April 1994; a total of 1.267 observations. The data are obtained from FRED database.

the estimated values along with robust standard errors of model (1). In particular, the model estimated by linear filter with Kim's approximation is labelled with Approximation 1. The following approximations instead are those in Section 5, respectively. Note that Approximation 2 mimics the auxiliary model of Gray (1996) and values are in fact in line (see Gray (1996), Table 3, p.44). The high-volatility regime is characterized by more sentivity to recent shocks ($\alpha_2 > \alpha_1$) and less persistence ($\beta_2 < \beta_1$) than the low-volatility regime. Within each regime, the GARCH processes are stationary ($\alpha_i + \beta_i < 1$) and the parameter estimates suggest that the regimes are very persistent, so the source of volatility persistence will be important. With regards to the MS-SV model, the approximated filters are presented in Section 6. Most of the masses in the transition probability matrix are concentrated in the diagonal, implying medium-high persistence in each regime. Moreover, the first regime is associated with a intermediate level of persistence in the volatility process while the second shows a highly-persistent volatility, with values close to one. In both models, however, the four approximations are not very dissimilar to the others.

Figure 2 contains plots of smoothed probabilities $Pr(s_t = 1|\Phi_T)$ which are of interest to determine if and when the regime switching occurs. The smoothed probability plots manage to identify crises periods that affected the market indices. The top panel of Figure 2 refers to the MS-GARCH and the bottom panel to the MS-SV model. In particular,

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
	(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
	(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
	(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
	(0.1140)	(0.0157)	(0.1629)	(0.2618)	(0.0642)	(0.1092)	(0.3603)	(0.2867)

Table 2: Estimation of the parameters in model (1) MS(2)-GARCH. Robust standard errors in parenthesis. 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.

Approximations	$\hat{p}$	$\hat{q}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\sigma}_v$	$\hat{\mu}_1$	$\hat{\mu}_2$
Approximation 1	0.8935	0.9542	0.5492	0.9256	0.2863	0.2460	0.4165
	(0.0345)	(0.3633)	(0.1252)	(0.1260)	(0.0208)	(2.5734)	(2.5797)
Approximation 2	0.8677	0.8676	0.5388	0.9852	0.3478	0.2952	0.4905
	(0.0347)	(0.2813)	(0.0651)	(0.0585)	(0.1148)	(1.3258)	(1.2977)
Approximation 3	0.8668	0.8435	0.6063	0.9730	0.2706	0.3143	0.5707
	(0.0197)	(0.2288)	(0.0863)	(0.0620)	(0.3524)	(1.7598)	(2.4015)
Approximation 4	0.8668	0.8435	0.6063	0.9730	0.2706	0.3143	0.5707
	(0.0197)	(0.2288)	(0.0863)	(0.0620)	(0.3524)	(1.7598)	(2.4015)

Table 3: Estimation of the parameters in model (11) MS(2)-SV. Robust standard errors in parenthesis. 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.

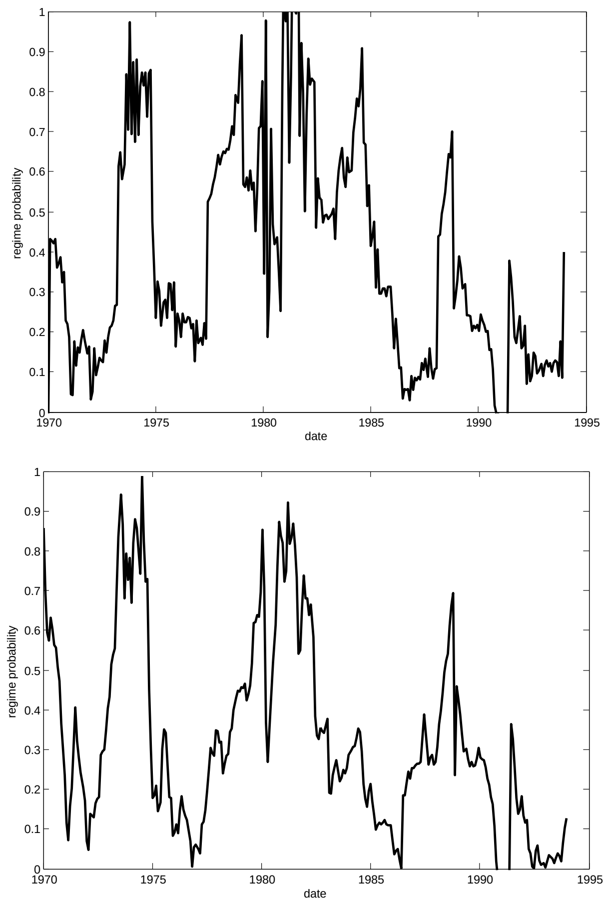

Figure 2: The top panel refers to MS-GARCH model and bottom panel to the MS-SV model. They represent smoothed probabilities being in high-volatility regime. Parameters estimates are based on a data set of one-month Treasury Bill rates, reported in annualized percentage terms. The sample period is from January 1970 to April 1994; a total of 1.267 observations. The data are obtained from FRED database.

both plots identify three periods of high-variance. The first (1973-1975) corresponds to the OPEC oil crisis. The second is shorter and more precise in the bottom panel and correspond to the Fed experiment (1979-1983). The third is a short period around 1987 after stock market crash.

8. Conclusion

In this paper we deal with Markov Switching models for volatility. In particular, we firstly consider MS-GARCH models which are known to suffer of path-dependence, i.e., dependence of the entire path history of the data. This makes Quasi Maximum Likelihood procedure unfeasible to apply. Hence, some solutions to overcome this problem have been proposed in the literature and particularly through the estimation of auxiliary models that allow similar effects of the original MS-GARCH. However, rewriting the model in a suitable state space representation, we propose an approximated linear filter following the line of Kim and Nelson (1999) and then we are able to prove duality in the estimation by Kalman filter and auxiliary models. Moreover, we introduce a linear filter also for MS-SV model on which different conditioning sets in the approximation step yield more flexibility in the estimation. We apply those methods to a simulation study and Treasury bill rates (the same dataset as in Gray (1996)). These applications show the feasibility of the linear filter for both MS models. In particular, this method has the advantage of avoiding fine-tuning procedures implemented in most Bayesian estimation techniques. In fact, giving some initial conditions, the only duty of the researcher is to decide which approximation to adopt in the filtering procedure. So, the proposed methods have a large applicability in financial and economics exercises and potential applications are those dealing with time varying volatility.

References

Augustyniak, M., (2013), "Maximum Likelihood Estimation of the Markov-switching GARCH model", forthcoming in *Computational Statistics and Data Analysis.*

Bauwens, L., Preminger, A., and Rombouts, J. (2010), "Theory and Inference for a Markov Switching GARCH Model", *Econometrics Journal*, Vol.13 **(2)**, 218–244.

Bauwens, L., Dufays, A., and Rombouts, J. (2011), "Marginal Likelihood for Markov-switching and Change-Point GARCH", *CORE discussion paper*, 2011/13.

Billio, M. and Caporin, M. (2005), "Multivariate Markov Switching dynamic conditional correlation GARCH representations for contagion analysis", *Statistical Methods and Applications*, **14**, 145–161.

Billio, M., Casarin, R., Osuntuy, A., (2012), "Efficient Gibbs Sampling for Markov Switching GARCH Models", Working Paper N. 35 , Dept. of Economics, University Ca' Foscari of Venice.

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 Czech Mathematicians and Physicists, 49–54.

Billio, M., Monfort A., and Robert, C.P., (1998b), "The simulated likelihood ratio (SLR) method", Document de Travail du CREST 9828, Paris.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, **52**, 5–59.

Bollerslev, T. (1990), "Modelling the coherence in short-run nominal exchage rates: a multivariate generalized arch model", *Review of Economics and Statistics*, **72**, 498–505.

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

Carvalho, C.M. and Lopes, H.F. (2007), "Simulation-based sequential analysis of Markov switching stochastic volatility models", *Computational Statistics and Data Analysis*, Vol.51 **(9)**, 4526-4542.

Casarin, R., (2004), Bayesian Inference for Generalised Markov Switching Stochastic Volatility Models, Cahier du CEREMADE N. 0414, University Paris Dauphine.

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

Engle, R. (2002), "Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models", *Journal of Business and Economic Statistics*, Vol.20 **(3)**, 339–350.

Francq, C. and Zakoïan, J.M. (2001), "Stationarity of Multivariate Markov Switching ARMA Models", *Journal of Econometrics*, Vol.102, 339–364.

Francq, C. and Zakoïan, J.M. (2012), "Strict Stationarity and Estimation of Explosive and Stationary Generalized Autoregressive Conditional Heteroscedasticity Models", *Econometrica*, Vol.80 **(2)**, 821–861.

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

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

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

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

Harvey, A.C., Ruiz, E. and Shephard, N. (1994), " Multivariate stochastic variance models", *Review of Economic Studies* **61**, 247–264.

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.

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

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

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.

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.

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

Pelletier, D. (2006), "Regime switching for dynamic correlation", *Journal of Econometrics*, **131**, 445–473.

Shephard, N. (1996), "Statistical aspects of ARCH and stochastic volatility", in *Time Series Models with Econometric, Finance and Other Applications*, Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (eds.), Chapman and Hall, London, 1–677.

So, K.P., Lam, K. and Li, W.K. (1998), "A stochastic volatility model with Markov switching", *Journal of Business and Economic Statistics*, Vol.16 **(2)**, 244–253.

Appendix

Appendix A – Computation details of some Formulae

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 ϵ_t which depends on the current regime s_t and the past regimes, i.e, $f(\epsilon_t|s_1,\ldots,s_t,\Psi_{t-1})$. In fact,

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

where

$$\begin{aligned}
f(\epsilon_t|s_1,\ldots,s_{t-1},\Psi_{t-1}) &= \sum_{s_t=1}^{M} f(\epsilon_t|s_1,\ldots,s_t,\Psi_{t-1})p(s_t|s_{t-1},\Psi_{t-1})\\
&= \sum_{s_t=1}^{M} f(\epsilon_t|s_1,\ldots,s_t,\Psi_{t-1})\, p_{t|t-1,t-1}.
\end{aligned}$$

A2. Here we derive the approximation of Kim and Nelson's Filter applied to model in (8), which is Formula (9):

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

Appendix B – Stationarity Conditions

Let us consider the MS-GARCH model in (1). Then we have

$$\begin{aligned}
E(\epsilon_t^2) = E(\sigma_t^2) = E(E(\sigma_t^2|s_t)) &= \sum_{j=1}^M E(\sigma_t^2|s_t=j)p(s_t=j)\\
&= \sum_{j=1}^M \pi_j(\omega_j+\alpha_j E(\epsilon_{t-1}^2)+\beta_j E(\sigma_{t-1}^2))\\
&= \sum_{j=1}^M \pi_j\omega_j + \sum_{j=1}^M \pi_j(\alpha_j+\beta_j)E(\sigma_{t-1}^2).
\end{aligned}$$

For any $n \geq 1$, we have

$$E(\sigma_t^2) = a \sum_{i=0}^{n-1} b^i + b^n E(\sigma_{t-n}^2)$$

where $a = \sum_{j=1}^{M} \pi_j \omega_j$ and $b = \sum_{j=1}^{M} \pi_j(\alpha_j + \beta_j)$. This immediately implies that the MS-GARCH process in (1) is *covariance stationary* if and only if $b < 1$. Of course, if $\delta_j = \alpha_j + \beta_j < 1$, for every $j = 1, \ldots, M$, the above condition is satisfied. Conversely, if the MS-GARCH is covariance stationary, at least one of the regimes is covariance stationary. The above condition is sufficient but non necessary for strict stationarity. By iteration, we get

$$\begin{aligned}
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t}\epsilon_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2 \\
&= \omega_{s_t} + \sigma_{t-1}^2(\alpha_{s_t} u_{t-1}^2 + \beta_{s_t}) \\
&= \omega_{s_t} + [\omega_{s_{t-1}} + \sigma_{t-2}^2(\alpha_{s_{t-1}} u_{t-2}^2 + \beta_{s_{t-1}})](\alpha_{s_t} u_{t-1}^2 + \beta_{s_t}) \\
&\vdots \\
&= \omega_{s_t} + \sum_{k=1}^{\infty} \omega_{s_{t-k}} \prod_{i=1}^{k} (\alpha_{s_{t-i+1}} u_{t-i}^2 + \beta_{s_{t-i+1}}).
\end{aligned}$$

For every $n \geq 2$, define

$$\sigma_{t,n}^2 = \omega_{s_t} + \sum_{k=1}^{n-1} \omega_{s_{t-k}} \prod_{i=1}^{k} a_{s_{t-i+1}}(u_{t-i}^2)$$

where $a_{s_t}(x) = \alpha_{s_t} x^2 + \beta_{s_t}$. Now

$$\sum_{k=1}^{n-1} \log[\omega_{s_{t-k}} \prod_{i=1}^{k} a_{s_{t-i+1}}(u_{t-i}^2)] = \sum_{k=1}^{n-1} \{\log \omega_{s_{t-k}} + \sum_{i=1}^{k} \log a_{s_{t-i+1}}(u_{t-i}^2)\}$$

is monotone. Then the limit $n \to +\infty$ is finite whenever

$$E[\log(\alpha_{s_t} u_{t-1}^2 + \beta_{s_t})] < 0.$$

Here log denotes the natural logarithm as usual. But we have

$$E[\log(\alpha_{s_t} u_{t-1}^2 + \beta_{s_t})] = E[E[\log(\alpha_{s_t} u_{t-1}^2 + \beta_{s_t})|s_t]] = \sum_{j=1}^{M} \pi_j E[\log(\alpha_j u_{t-1}^2 + \beta_j)].$$

So we get that $\sigma_t^2 < +\infty$ a.s. (almost surely) and $\{\epsilon_t^2, \sigma_t^2\}$ is *strictly stationary* if

$$\sum_{j=1}^{M} \pi_j E[\log(\alpha_j u_{t-1}^2 + \beta_j)] < 0.$$

This extends the strictly stationarity condition given by Francq and Zakoïan (2012) for a GARCH(1,1) model to the case of changing in regime. See also Theorem 1 in Bauwens et al. (2010). Of course, the covariance stationarity condition implies strict stationarity, but the converse is not true in general.

The MS-GARCH(1,1) model in (1) can be represented by a MS-ARMA(1,1) process as in (7)

$$(1-\delta_{s_t}L)\epsilon_t^2 = \omega_{s_t} + (1+\theta_{s_t}L)v_t$$

where $\delta_{s_t} = \alpha_{s_t} + \beta_{s_t}$ and $\theta_{s_t} = -\beta_{s_t}$. The necessary and sufficient condition for second-order stationarity of univariate MS-ARMA(1,1) models was given by Francq and Zakoïan (2001), see Example 3 pag.351. We apply their result in our case. Let us consider the $M \times M$ matrix

$$\Omega = (a_{ij})_{i,j=1,\ldots,M}$$

where $a_{ij} = p_{ji}\delta_i^2$. Let $\rho(\Omega)$ be the spectral radius of the matrix, that is, its largest eigenvalue in modulus. From Francq and Zakoïan (2001), $\rho(\Omega) < 1$ if and only if the process (ϵ_t^2) in (1) is *second-order stationary* in the case where, for at least one regime, the AR and MA polynomials have no common roots. For our model, this means that $\delta_j \neq -\theta_j$, that is, $\alpha_j > 0$ for some $j = 1, \ldots, M$. Finally, note that the MA part in the process (ϵ_t^2) does not matter for the second-order stationarity condition.

Finally, with regards to the MS-SV model in (12), its MS-ARMA representation is easily obtained as follows

$$(1-\rho_{s_t}L)y_t = \xi_{s_t} + (1+\beta_{s_t}L)z_t$$

where $\xi_{s_t} = \alpha - \rho_{s_t}\alpha + \mu_{s_t}$ and $z_t + \beta_{s_t}z_{t-1} = v_t + e_t - \rho_{s_t}e_{t-1}$. Thus stationarity conditions as discussed above apply. More precisely, the process is second-order stationary if and only if $\rho(\widetilde{\Omega}) < 1$, where $\widetilde{\Omega}$ is the matrix obtained by replacing δ_j by ρ_j in the definition of Ω.