

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

# Computational Statistics and Data Analysis

journal homepage: [www.elsevier.com/locate/cstda](http://www.elsevier.com/locate/cstda)

## Long memory and nonlinearities in realized volatility: A Markov switching approach

Davide Raggi<sup>a,\*</sup>, Silvano Bordignon<sup>b</sup>

<sup>a</sup> University of Bologna, Department of Economics, Piazza Scaravilli 2, 40126 Bologna, Italy

<sup>b</sup> University of Padova, Department of Statistics, Via C. Battisti 231, 35123 Padova, Italy

### ARTICLE INFO

#### Article history:

Available online 29 December 2010

#### Keywords:

Realized volatility  
Switching-regime  
Long memory  
MCMC  
Forecasting

### ABSTRACT

Realized volatility is studied using nonlinear and highly persistent dynamics. In particular, a model is proposed that simultaneously captures long memory and nonlinearities in which level and persistence shift through a Markov switching dynamics. Inference is based on an efficient Markov chain Monte Carlo (MCMC) algorithm that is used to estimate parameters, latent process and predictive densities. The in-sample results show that both long memory and nonlinearities are significant and improve the description of the data. The out-sample results at several forecast horizons show that introducing these nonlinearities produces superior forecasts over those obtained using nested models.

© 2010 Elsevier B.V. All rights reserved.

### 1. Introduction

It is well-known that accurately measuring and forecasting financial volatility plays a central role in many pricing and risk management problems. With high frequency intra-daily data sets becoming widely available, more accurate estimates of volatility can be obtained. Realized volatility (RV), i.e. the sum of intra-day squared returns, reduces the noise in the volatility estimate considerably when compared to other volatility measures, such as squared or absolute daily returns. Thus, volatility becomes *observable* and can be modeled directly, rather than being treated as a latent process both in a GARCH or in a stochastic volatility setup.

Many empirical regularities on RV have been well-documented in recent literature and a detailed review has been provided by McAleer and Medeiros (2008b). One of the most relevant empirical regularities is that RV exhibits high persistence, as is evidenced in the research of Andersen et al. (2003), Koopman et al. (2005) and Corsi (2009). For this reason, linear fractionally integrated models (ARFIMA) are generally used to capture this feature. A flexible strategy to model serial dependencies for RV has been proposed in Barndorff-Nielsen and Shephard (2002) through a superposition of ARMA(1, 1) processes, whereas Corsi (2009) has introduced the Heterogeneous Autoregressive model (HAR) given by a combination of volatilities measured over different time horizons. However, Granger and Ding (1996) found out that persistence in volatility tends to be non-constant over time and Longin (1997) provided evidence of an usually high level of persistence when volatility is low, thus suggesting the presence of nonlinearities.

On the other hand, it is well-known that a long memory can be overestimated when a regime shifts or structural breaks are not taken into account. In fact, these are confounding factors and distinguishing between them can be rather troublesome (Diebold and Inoue, 2001). However, recent statistical tests aimed at disentangling the effects of long memory and level shifts have been proposed in Baillie and Kapetanios (2007) and in Ohanissian et al. (2008). Baillie and Kapetanios (2007) found the presence of nonlinearity together with a long memory in realized volatilities for currencies, whereas Ohanissian et al. (2008), in their empirical application, provided evidence that realized volatility of exchange rates, such as DM/\$ and

\* Corresponding author.

E-mail addresses: [davide.raggi@unibo.it](mailto:davide.raggi@unibo.it) (D. Raggi), [bordigno@stat.unipd.it](mailto:bordigno@stat.unipd.it) (S. Bordignon).

Yen/\$ is properly described by a true long memory process. This latter result seems in contrast with Perron and Qu (2010), who claim that short memory models with level shifts are appropriate to describe volatility. Similarly, Carvalho and Lopes (2007) model volatility with a short memory switching regime dynamics. Chen et al. (2008) propose a range-based threshold heteroskedastic model, whereas He and Maheu (2010) propose a GARCH model subject to an unknown number of structural breaks.

The question of whether long memory is spurious is still debateable. It thus appears of obvious interest to join both models' features into a single time series model. In this way, it is also possible to check whether the benefits of combining long memory and nonlinearities represent an improvement in forecasting accuracy. Recently, some time series models have suggested combining long memory and nonlinearities to describe conditional variances. Contributions in this direction are, respectively, Martens et al. (2004) and Hillebrand and Medeiros (2008) who built from an ARFIMA model by allowing for smooth level shifts, day of the week effects and leverage. McAleer and Medeiros (2008a) propose a different strategy by introducing a multiple regime smooth transition extension of HAR. Finally, Lux and Morales-Arias (2010) introduce a regime-switching multifractal model.

In this paper, we consider a different strategy to model abrupt changes in the conditional mean and time varying long range dependency. We base our analysis on a Markov switching model that is in line with the seminal work of Hamilton (1989) in which level shifts are modeled through a binary non-observable Markov process and in which the parameters, including the degree of persistence, are state dependent. Persistence is introduced through a standard ARFIMA model.

Furthermore, we think it is important to include exogenous regressors in the model's specification. Following Bandi and Perron (2006), we consider the implied volatility as a predictor, since it is proven to be an unbiased long run forecast of future RV, once controlling for a fractionally cointegration relation. An interesting study on the relation between realized volatility, long range persistence and exogenous regressors has recently been completed by Corsi and Renò (2010).

We use Bayesian estimation techniques and goodness-of-fit indicators to assess the in-sample performance of our model. We also pay particular attention to the forecasting ability of the proposed models.

We base our empirical analysis on the 5 min intra-daily series of Standard & Poor's 500 (S&P500) stock index over the period 1 January 2000 to 28 February 2005. Our results showed that implied volatility is important for predicting RV and also that, in the short run, long memory, together with nonlinearities improve forecasting performance. In the long run, the ARFIMA effect seems to be dominant with respect to the switching regime mechanism.

The remainder of the paper is organized as follows. Our data set is described in Section 2. Section 3 contains the definitions of the Markov switching models that are considered in this research. Our inferential solution is outlined in Section 4. The forecasting methodology is explained in Section 5 and empirical results based on simulated and real data are illustrated in Section 6 and in Section 7, respectively.

## 2. Realized and implied volatility

Realized volatility is an efficient and unbiased measure of the actual volatility based on the quadratic variation of a stochastic process. Theoretical and empirical features on this subject have been investigated in Barndorff-Nielsen and Shephard (2002) and in Andersen et al. (2003). Consider, for instance, a simple continuous time model for the log-price of a financial security

$$p(t + \tau) = \mu(t + \tau)d\tau + \sigma(t + \tau)dW(t + \tau) \quad 0 \leq \tau \leq 1, t = 1, 2, \dots$$

in which  $W(t)$  is a standard Brownian motion. Using well-known results on stochastic processes, it can be proved that daily returns,  $r_t = p(t) - p(t - 1)$ , are Gaussian with conditional distribution

$$r_t | \mathcal{F}_t \sim \mathcal{N} \left( \int_0^1 \mu(t + \tau - 1)d\tau, \int_0^1 \sigma^2(t + \tau - 1)d\tau \right),$$

where the variance is known as the Integrated Volatility at day  $t$ , namely,  $IV_t$ . It can be proved that, in the absence of microstructure noise, i.e., the observed prices are not affected by measurement errors, IV can be consistently estimated by the realized volatility and defined as

$$\hat{y}_t = RV_t = \sum_{j=1}^{N_t} [p_{j,t} - p_{j-1,t}]^2, \quad j = 1, \dots, N_t, t = 1, \dots, T \tag{1}$$

in which  $p_{j,t}$  is the  $j$ -th observation at day  $t$  and  $N_t$  is the number of intra-daily observations. In the case of microstructure noise, the true prices  $p_{j,t}^*$  are affected by measurement errors, leading to  $p_{j,t} = p_{j,t}^* + \epsilon_{j,t}$ . As a main consequence, the realized volatility estimator diverges when data are sampled too frequently over the day (Bandi and Russell, 2008). However, as suggested in Andersen and Bollerslev (1998) and in Andersen et al. (2001), in practical applications the realized volatility is still a valid estimator when data are sampled at a lower frequency with respect to the tick-by-tick, since the microstructure noise becomes negligible and then a sampling frequency ranging between 5 to 30 min can be a reasonable choice. In our study, we also consider the CBOE Volatility Index (VIX), which is a measure of the market's expectation of 30-day volatility implied by at-the-money S&P500 Index option prices provided by the Chicago Board Options Exchange. It is computed by averaging the weighted prices of S&P500 puts and calls over a wide range of strike prices. Since VIX refers to option contracts, it can be seen as market predictor of expected volatility.

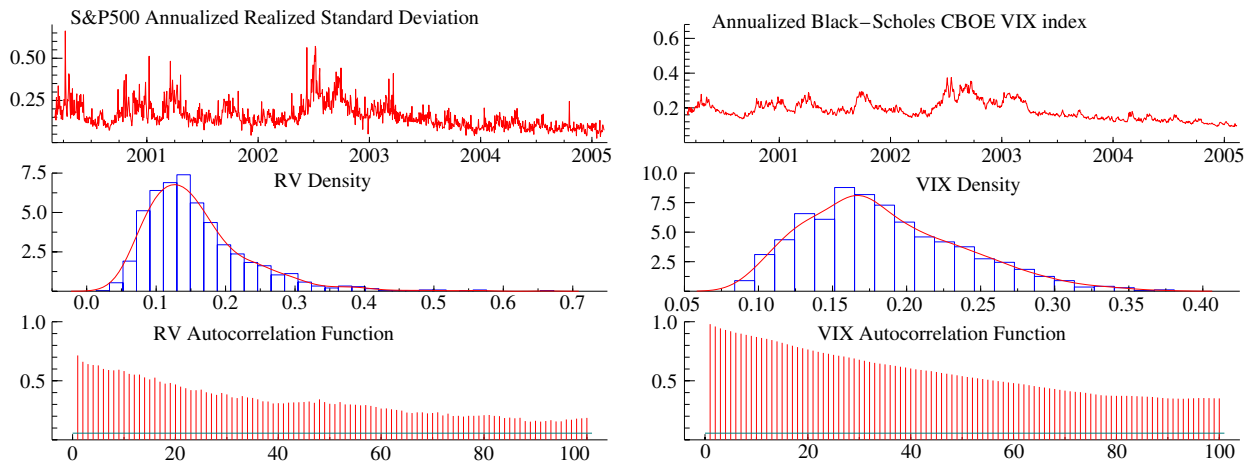


Fig. 1. Realized standard deviations and VIX index data from 01/03/2000 to 28/02/2005, estimated densities and autocorrelation function.

Table 1

Descriptive statistics of RV and VIX. We also report the estimated  $d$  of the ARFIMA(0,  $d$ , 0) model obtained through exact maximum likelihood and computed using the Arfima package of Doornik and Ooms (2006). Standard errors are in parentheses.

	Realized volatility	Implied volatility (VIX)
Mean	0.15891	0.18512
Median	0.14255	0.17632
Std. Dev.	0.076078	0.052924
Skewness	1.6990	0.66155
Kurtosis	4.8575	0.086316
Minimum	0.025933	0.092231
Maximum	0.66168	0.37457
$d$	0.381337 (0.01973)	0.499673 (0.00041)

The S&P500 index data set, provided by Olsen and Associates in Zurich, consists of 5 min intra-daily observations ( $N_t = 288$ ), whereas VIX, downloaded from the CBOE website, is observed on a daily basis from January 1, 2000 to February 28, 2005. We have removed from our sample days in that the market has been closed, such as the three weeks after September 11, weekends and US holidays. This resulted in a total number of  $T = 1274$  trading days. Following Bandi and Perron (2006), we consider the annualized realized standard deviation  $y_t = \frac{\sqrt{252 \times \hat{y}_t}}{100}$  and the annualized VIX which is also multiplied by  $\sqrt{\frac{252}{365}}$  in order to account for the difference between trading days and calendar days in a year. Descriptive statistics are reported in Table 1; Fig. 1 displays the dynamics of the volatility measures. In particular, from Fig. 1, we observe high volatility levels at the beginning of 2000, from October 2000 until March 2001 and from April 2002 to March 2003. From Table 1, it is clear that both time series are highly persistent since the Fractional Integration parameter  $d$  is always larger than 0.35.

### 3. A switching regime model for long memory realized volatility

We propose a long memory switching regime process based on a mixture of two ARFIMA(0,  $d_i$ , 0),  $i = 0, 1$  dynamics and hypothesize difference persistence levels according to two volatility regimes. We label the regime with low volatility and small conditional variance, which can be interpreted as the stable and normal volatility state of the market, as 0, whereas label 1 indicates high and fluctuating volatility. Two regimes models have been proposed in Cai (1994), Dueker (1997), Susmel (2000), Klaassen (2002) and Marcucci (2005). Generalization for more regimes, thus allowing for low, moderate and high volatility, has been proposed, for instance by Hamilton and Susmel (1994), Haas et al. (2004), Kauffmann and Frühwirth-Schnatter (2002) and Ardia (2009). Unfortunately, building tests to detect the number of regimes is a difficult task. For instance, Kauffmann and Frühwirth-Schnatter (2002)'s show that a formal test within a classical framework of maximum likelihood is not possible, since the regularity conditions for justifying the  $\chi^2$  approximation to the likelihood ratio statistics do not hold. Some difficulties also arise in a Bayesian setup because of the sensitivity of the results on the prior specification of the switching parameters. We thus consider a two regimes model, first because, to our knowledge, is the most popular choice in the literature, and second because there is no strong evidence that suggests the need of more than two regimes. In a private equity market context, Susmel (2000) has evidenced that a third regime has been detected in correspondence with the crash of October 1987. In particular, the introduction of a third regime provided a negligible improvement of the likelihood with respect to the two states alternative, even though the general specification appeared to be useful for dating volatility states. Similar findings have been obtained in Henneke et al. (2009) in which the third regime

appears just once in their empirical analysis thus casting doubt on the need to add an extra state and suggesting the presence of a structural break. Finally, [Ardia \(2009\)](#), in a more general setup, claimed that a three state Markov switching model is not clearly supported by the data.

Our specification is similar to [Ray and Tsay \(2002\)](#), in which, first, ARFIMA models with random level shifts but constant persistence are introduced and second, ARFIMA dynamics with time varying persistence levels but regime invariant parameters are proposed. Here we consider the regime switching Markov model of [Hamilton \(1989\)](#) with state dependent fractionally integration parameters, labeled as *Model 1*, that allow for level shifts together with time varying persistence

$$y_t = \mu_{S_t} + \beta_{S_t} X_t + (1 - L)^{-d_{S_t}} \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2), \tag{2}$$

in which  $y_t$ ,  $t = 1, \dots, T$  are the realized volatilities and  $S_t$ ,  $t = 1, \dots, T$  is an unobservable two states  $\{0, 1\}$  first order Markov chain with transition probability matrix  $\Pi = \{\pi_{ij}\}$ ,  $i, j = 0, 1$ . In our exercise,  $X_t$ ,  $t = 1, \dots, T$  is the implied volatility index VIX. To avoid some potential endogeneity troubles, we consider as  $X_t$  the delayed regressor, i.e.,  $VIX_{t-1}$ . This parameterisation encompasses many other special cases, and, in particular we consider a model with common  $\beta$  and persistence but regime-dependent levels and conditional variances, namely *Model 1A*

$$y_t = \mu_{S_t} + \beta X_t + (1 - L)^{-d} \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2), \tag{3}$$

and the non-switching ARFIMA(0,  $d$ , 0) with VIX, that is *Model 1B*,

$$y_t = \mu + \beta X_t + (1 - L)^{-d} \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{4}$$

We also consider *Model 1C* that is a short memory switching regime model with VIX

$$y_t = \mu_{S_t} + \beta_{S_t} X_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2). \tag{5}$$

Special cases obtained by omitting the exogenous regressor are also taken into account and respectively labeled as *Model 2*, *Model 2A* and *Model 2B*. Finally *Model 3* is the pure Markov switching model

$$y_t = \mu_{S_t} + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2). \tag{6}$$

To complete the model's specification, we consider the following priors:  $\mu_i \sim \mathcal{N}(1, 0.5)$ ,  $\beta_i \sim \mathcal{N}(1, 0.5)$ ,  $d_i \sim \mathcal{U}(0, 0.5)$ ,  $\sigma_i^2 \sim IG(2.5, 0.75)$  and finally  $p_{ij} \sim \text{Beta}(5, 95)$ ,  $i, j = 0, 1$ . The priors for  $\mu_i$  and  $\beta_i$  are centered respectively at 0 and 1 since [Bandi and Perron \(2006\)](#) showed that VIX is an unbiased predictor of the realized volatility. Furthermore, an uninformative prior on  $d_i$  reflects our a priori lack of knowledge on these parameters. It is worth noting that we consider state independent priors as suggested in [Frühwirth-Schnatter \(2001\)](#).

#### 4. MCMC methodology

Our goal is to jointly estimate long memory dynamics and the latent process. Inference for this model is not obvious, since the latent process  $\mathbf{S} = (S_1, \dots, S_T)$  is not observable and also because inference for long range parameters is not standard. Inference for regime switching models can be performed in the classical, as well as in the Bayesian framework, and many inferential techniques have been proposed to numerically evaluate the likelihood function. A detailed review on this topic can be found in [Kim and Nelson \(1999\)](#) and in [Frühwirth-Schnatter \(2006\)](#). Furthermore, a unique estimator is not available for  $d_i$ ,  $i = 0, 1$  and a detailed review on inference for long memory parameters has been provided by [Palma \(2007\)](#). Bayesian estimation for ARFIMA model has been taken into account in a number of papers. [Koop et al. \(1997\)](#) derive predictive and posterior densities through an importance sampling integration scheme and provide a way to deal with model uncertainty. MCMC sampling schemes have been proposed in [Pai and Ravishanker \(1996, 1998\)](#) and [Hsu and Breidt \(2003\)](#). They basically differ in the way the likelihood function is computed or approximated, whereas [Brockwell \(2006\)](#) exploits an efficient sampling scheme for a general class of long range dependent models. A different approach, based on wavelets, has been implemented in [Ko and Vannucci \(2006\)](#). Finally, [Chen and Yu \(2005\)](#) propose a hybrid MCMC scheme to estimate a long memory model for stock returns with asymmetric GARCH errors.

In this paper, we focus on MCMC methods for inference, since they allow one to take into account long memory and regime-switching jointly. In particular, the state space representation of an ARFIMA model can be easily extended to the case of time varying parameters.

Our solution is based on [Chan and Palma \(1998\)](#), in which it is proved that the exact likelihood of an ARFIMA process can be recursively computed by means of the Kalman filter in a finite number of steps, even though the system has an infinite dimensional representation. For instance, the ARFIMA(0, $d$ ,0) model has a linear MA representation given by

$$y_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \epsilon_{t-j} \quad \epsilon_t \sim WN(0, \sigma^2). \tag{7}$$

To make the Kalman recursions feasible, [Chan and Palma \(1998\)](#) consider an approximation of Eq. (7) based on a truncation of order  $M$  that can be written in state space form. [Ray and Tsay \(2002\)](#) exploited this representation to derive inference



for long memory processes with random level shifts in an MCMC setup. It is worth noting that the ARFIMA dynamics can also be written through an autoregressive representation, even though Palma (2007) suggests that an MA approximation guarantees computational advantages.

In a switching regime setup, conditionally on the regimes vector  $\mathbf{S}$ , we can represent the dynamics defined in (2) by the time varying parameter state space system given as follows

$$\underbrace{\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \\ \vdots \\ \epsilon_{t-M+1} \end{bmatrix}}_{\alpha_{t+1(M+1) \times 1}} = \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}}_{T_{(M+1) \times (M+1)}} \underbrace{\begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-M} \end{bmatrix}}_{\alpha_t} + \underbrace{\begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{u_{t+1(M+1) \times 1}} \quad (8)$$

$$y_t = \mu_{S_t} + \beta_{S_t} X_t + \underbrace{\begin{bmatrix} 1 & \pi_1(d_{S_t}) & \pi_2(d_{S_t}) & \dots & \pi_M(d_{S_t}) \end{bmatrix}}_{Z_t \ 1 \times (M+1)} \alpha_t, \quad (9)$$

in which  $\pi_j(d_{S_t}) = \frac{\Gamma(j+d_{S_t})}{\Gamma(j+1)\Gamma(d_{S_t})}$ . Inference for the parameters can thus be derived by applying standard results on MCMC for ARMA models for which seminal contributions are Albert and Chib (1993), McCulloch and Tsay (1993), Billio et al. (1999) and Kim and Nelson (1999). However, to our knowledge, MCMC methodologies have never been implemented for ARFIMA model with switching regime parameters.

We propose an MCMC algorithm to simulate the posterior distribution of  $(\theta, \mathbf{S})$ , where  $\theta = (\mu_i, \beta_i, \sigma_i, d_i, p_{01}, p_{10}) \ i = 0, 1$ , is the parameter's vector. Furthermore, since our procedure is based on an approximated model, we consider the truncation parameter  $M$  as unknown. Thus we build an algorithm that selects the appropriate level of truncation.

In the follow, we will call  $\theta_j$  the generic  $j$ -th element of  $\theta$ . Finally  $\mathbf{Y}$  contains the observable realized volatilities, whereas  $\mathbf{X}$  holds the exogenous regressor's realization.

The basic idea behind MCMC is to simulate a trajectory of a Markov chain  $\{M^{(j)}, \theta^{(j)}, \mathbf{S}^{(j)}\}_{j=1}^n$  from a given starting point  $(M^{(0)}, \theta^{(0)}, \mathbf{S}^{(0)})$ , with limiting invariant distribution  $p(M, \mathbf{S}, \theta | \mathbf{Y}, \mathbf{X})$ . See Robert and Casella (1999, ch. 6–7) for a general treatment on the conditions needed to achieve convergence of MCMC algorithms. Once convergence is achieved, the algorithm provides a sample of serially dependent draws for  $M, \theta$  and  $\mathbf{S}$ , which can be used to perform inference. More precisely, estimates of the latent factors and of the posterior mean of  $\theta$  are given by averaging over the realization of the chain. Moving the whole vector  $(M, \theta, \mathbf{S})$  in a block can be difficult, since it is highly multivariate. A possible strategy is to divide it into sub-components and then update them individually. As suggested in Shephard (1994) and Carter and Kohn (1994), updating the whole latent process  $\mathbf{S}$  in a block from its joint conditional distribution should reduce the autocorrelation between states and then speed up the convergence of the chain to its invariant distribution. Here the regime switches are updated through the multi-move Gibbs sampler of Chib (1996), whereas parameters are updated individually. We summarize the algorithm as follows

- Initialize the chain at  $(M^{(0)}, \theta^{(0)}, \mathbf{S}^{(0)})$
- At step  $j = 1, \dots, n$ 
  - Update  $M$  following Troughton and Godsill (1998);
  - Update  $\theta$  one-at-a-time from  $p(\theta_i | M^{(j)}, \mathbf{S}^{(j-1)}, \theta_{-i}^{(j)}, \theta_{-i+}^{(j-1)}, \mathbf{Y}, \mathbf{X})$  through Metropolis–Hastings, where  $\theta_{-i-}$  are the first  $(i - 1)$  elements of  $\theta$  and  $\theta_{-i+}$  includes the  $(i + 1)$ -th to the last;
  - Draw  $\mathbf{S}^{(j)}$  in a block from  $p(\mathbf{S} | M^{(j)}, \theta^{(j)}, \mathbf{Y}, \mathbf{X})$ , using Chib (1996).

Details of this scheme are described in the next section.

#### 4.1. Choosing the truncation lag $M$

The problem of choosing the truncation parameter is similar in principle to the selection of the model order for ARMA processes. In a Bayesian framework, this issue has been tackled through the Reversible Jump MCMC algorithm (RJMCMC) of Green (1995). Efficient versions of the RJMCMC for autoregressive dynamics have been proposed in Troughton and Godsill (1998) and in Vermaak et al. (2004), whereas a generalization to ARMA models is due to Ehlers and Brooks (2004). Here the model order uncertainty issue is simpler, since the dimension of the parameter space remains constant when  $M$  changes. Thus, our procedure is a special case of Troughton and Godsill (1998). We propose a new truncation parameter  $M^*$  from a discretized Laplace proposal  $q(M | M^{(j-1)}) \propto \exp\{-\lambda |M - M^{(j-1)}|\}$ , in which  $\lambda = 0.1$  usually to obtain small moves with occasional large jumps. The acceptance probability becomes

$$\alpha(M^{(j-1)}, M^*) = \frac{p(M^* | \mathbf{Y}, \mathbf{X}, \mathbf{S}, \theta) q(M^{(j-1)} | M^*)}{p(M^{(j-1)} | \mathbf{Y}, \mathbf{X}, \mathbf{S}, \theta) q(M^* | M^{(j-1)})}$$

As suggested in Troughton and Godsill (1998), a suitable prior for  $M$  is the Poisson truncated distribution with  $M \in \{M_{\min} \dots, M_{\max}\}$  that, in our application, is fixed to 10 and 50, respectively.

#### 4.2. Updating the parameters

It is easy to show that the conditional posterior of  $p_{ij}$ , and  $\sigma_i^2$ ,  $i = 0, 1$  are respectively Beta and Inverse Gamma. Furthermore, Billio et al. (1999) provided a closed form expression for the full conditional of  $\mu_i$ ,  $i = 0, 1$ . Sampling  $\beta_i$  and  $d_i$ ,  $i = 0, 1$  is more involved, since their conditional distributions are not known in closed form. We propose to update each of these parameters in turn using the Metropolis–Hastings algorithm. At the  $j$ -th step of the algorithm, we simulate  $\theta_i$  as follows

- Sample  $\theta_i^*$  from a random walk proposal  $q(\theta_i | \theta_i^{(j-1)})$ .
- Define  $\alpha(\theta_i^{(j-1)}, \theta_i^*) = \min \left( \frac{p(\theta_i^*)p(\mathbf{Y}|M^{(j)}, \theta_i^*, \mathbf{X}, \mathbf{S}^{(j)}, \theta_{-i}^{(j)}, \theta_{-i+}^{(j-1)})q(\theta_i^{(j-1)} | \theta_i^*)}{p(\theta_i^{(j-1)})p(\mathbf{Y}|M^{(j)}, \theta_i^{(j-1)}, \mathbf{X}, \mathbf{S}^{(j)}, \theta_{-i}^{(j)}, \theta_{-i+}^{(j-1)})q(\theta_i^* | \theta_i^{(j-1)})}, 1 \right)$ .
- Sample  $u$  from  $U(0, 1)$ . If  $u \leq \alpha(\theta_i^{(j-1)}, \theta_i^*)$ 
  - then  $\theta_i^{(j)} = \theta_i^*$
  - otherwise  $\theta_i^{(j)} = \theta_i^{(j-1)}$ .

Here  $p(\theta_i)$  is the prior distribution for the  $i$ -th parameter and  $p(\mathbf{Y}|M, \theta, \mathbf{X}, \mathbf{S})$  is the likelihood of the model given  $\mathbf{S}$ ,  $M$  and  $\theta$ . Evaluating  $p(\mathbf{Y}|M, \theta, \mathbf{X}, \mathbf{S})$  is not trivial, however it can be done via the Kalman filter.

To identify the two states, a common strategy is to impose some constraints in the parameters' space. Here we recur to the systematic permutation sampler proposed in Frühwirth-Schnatter (2001) in which, first, all the quantities of interest are sampled from their unconstrained posterior and, second, a permutation of the current labeling of the states of the latent process applies in case the identification constraint is not satisfied. A preliminary run of the unconstrained MCMC sampler on real data evidenced the presence of two distinct regimes that differ in the intercepts and in the variances. In particular, a low intercept corresponds to a small variance. For this reason,  $\sigma_0^2 < \sigma_1^2$  or  $\mu_0 < \mu_1$  provide exactly the same results. In our experience, the output of the permutation sampler based on  $\sigma_0^2 < \sigma_1^2$  did not show any sign of label switching, thus suggesting that the constraint actually separates the two regimes.

#### 4.3. Updating S

We update  $\mathbf{S}$  in a block, moving the vector according to the algorithm proposed in Chib (1996). The method is based on the following decomposition of  $p(\mathbf{S}|M, \mathbf{Y}, \mathbf{X}, \theta)$ , in which we drop the dependence on  $M$ ,  $\theta$  and  $\mathbf{X}$  for notational convenience

$$p(\mathbf{S}|\mathbf{Y}) = p(S_T|\mathbf{Y}) \dots p(S_t|\mathbf{Y}, S_{t+1}, \dots, S_T) \dots p(S_1|\mathbf{Y}, S_2, \dots, S_T),$$

where, in particular

$$p(S_t|\mathbf{Y}, S_{t+1}, \dots, S_T) \propto p(S_t|y_1, \dots, y_{t-1})p(y_t|y_1, \dots, y_{t-1})p(S_{t+1}|S_t),$$

in which  $p(S_t|y_1, \dots, y_{t-1}) = \sum_{j=0}^1 p(S_t|S_{t-1} = j)p(S_{t-1} = j|y_1, \dots, y_{t-1})$  and  $p(S_t|y_1, \dots, y_t) \propto p(S_t|y_1, \dots, y_{t-1})p(y_t|y_1, \dots, y_{t-1})$ . Numerical evaluation of  $p(y_t|y_1, \dots, y_{t-1})$  is provided by applying the Kalman filter to ((8)–(9)), whereas  $p(S_t|y_1, \dots, y_{t-1})$  and  $p(S_t|y_1, \dots, y_t)$  can be computed recursively starting from  $p(S_1|y_0) = p(S_1)$ , i.e., the stationary distribution of the Markov chain. Once all of these quantities are computed, it is easy to draw  $S_T$  from  $p(S_T|\mathbf{Y})$  that is binomial, while  $S_{T-1}, \dots, S_1$  can be recursively obtained from  $p(S_t|\mathbf{Y}, S_{t+1}, \dots, S_T)$ .

#### 4.4. Forecasting

MCMC allows for the easy approximation of the density forecasts of  $y_{T+1}, \dots, y_{T+h}$  for a given horizon  $h > 0$ . At each Gibbs iteration, we can simulate future trajectories of latent and observed processes, i.e.,  $y_{T+1}, \dots, y_{T+h}$  and  $S_{T+1}, \dots, S_{T+h}$  given  $\theta, \mathbf{X}$  and  $M$ . To simplify the notation, we define  $\mathbf{X}_h = (X_{T+1}, \dots, X_{T+h})$ ,  $\mathbf{S}_h = (S_{T+1}, \dots, S_{T+h})$  and  $\mathbf{Y}_h = (y_{T+1}, \dots, y_{T+h})$  respectively. A sample from the predicting densities can be simulated as:

- At the  $j$ -th Gibbs iteration set  $\theta = \theta^{(j)}$  and  $M = M^{(j)}$ ;
  - Draw  $S_{T+i}^{(j)}$  from  $p(S_{T+i}|S_{T+i-1}^{(j)}, \theta^{(j)})$ , for  $i = 1, \dots, h$ ;
  - Draw  $y_{T+i}^{(j)}$  from  $p(y_{T+i}|\mathbf{Y}, \mathbf{Y}_{T+i-1}^{(j)}, M, \mathbf{X}_{T+i}, \mathbf{S}, S_{T+i}^{(j)}, \theta)$ ,  $i = 1, \dots, h$ .

### 5. In- and out-of-sample comparisons

To compare the out-of-sample predictive accuracy of the models, we use the *Weighted Likelihood Ratio test (WLR)* of Amisano and Giacomini (2007), which is based on density forecasts computed through classic or Bayesian techniques. WLR tests require a rolling window estimation procedure, that is, the recursion of the  $h$ -steps-ahead,  $h \geq 1$ , density forecasts have to be obtained using a fixed sample of size  $N$ . In practice, the first  $h$ -steps-ahead forecast is based on  $(y_1, \dots, y_N)$ , the second on  $(y_2, \dots, y_{N+1})$ , until the last one, that is based on  $(y_{T-N-h+1}, \dots, y_{T-h})$ . We thus obtain a sequence of  $K_h = T - N - h - 1$

density forecasts. The comparison between two models, namely  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , is based on the weighted difference of two loss functions called *scoring rules*, defined as  $L_i(y_{t+h}) = f_i(y_{t+h})$  and  $L_j(y_{t+h}) = f_j(y_{t+h})$ , in which  $f_i(\cdot)$  and  $f_j(\cdot)$  are the  $h$ -steps-ahead density forecasts for  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , whereas  $y_{t+h}$  is the actual future observation. The test is thus based on

$$WLR_{t+h} = \omega(\tilde{y}_{t+h})[\log \hat{f}_i(y_{t+h}) - \log \hat{f}_j(y_{t+h})]$$

where  $\tilde{y}_{t+h}$  is the standardized observation,  $\hat{f}_i$  and  $\hat{f}_j$  are the predictive densities estimated by means of  $(y_{t-N+1}, \dots, y_t)$  under models  $\mathcal{M}_i$  and  $\mathcal{M}_j$  and finally  $\omega(\cdot)$  is a weight that allows the emphasis of some specific region of the support of the density forecasts. In the MCMC setup, these predictive densities can be easily estimated as described in Section 4. Note that a positive difference means a superior predictive accuracy of model  $i$  versus model  $j$ . Following Amisano and Giacomini (2007), a test for equal performance of  $h$ -steps-ahead density forecasts  $f_i(\cdot)$  and  $f_j(\cdot)$  can be formulated as a test of the hypothesis system

$$H_0 : E[WLR_{t+h}] = 0, t > N \quad \text{vs.} \quad H_1 : E \left[ \sum_{t=N}^{T-h} \frac{WLR_{t+h}}{K_h} \right] \neq 0,$$

and the associated statistic is

$$t_h = \frac{\frac{1}{K_h} \sum_{t=N}^{T-h} WLR_{t+h}}{\widehat{\text{avar}}(WLR_{t+h})},$$

where  $\widehat{\text{avar}}(WLR_{t+h})$  is a consistent estimate of the long range asymptotic variance of the numerator that can be estimated by using estimators robust to heteroskedasticity and autocorrelation (see Newey and West, 1987, for instance). Under technical conditions, it can be proved that  $t_h \stackrel{a}{\sim} \mathcal{N}(0, 1)$ . We also consider the Root Mean Square Error (RMSE) together with the Mean Absolute Error (MAE), that is, for model  $i$   $\text{RMSE}_i = \sqrt{\frac{1}{K_h} \sum_{t=N}^{T-h} \epsilon_{t+h|t,i}^2}$  and  $\text{MAE}_i = \frac{1}{K_h} \sum_{t=N}^{T-h} |\epsilon_{t+h|t,i}|$ , where  $\epsilon_{t+h|t,i} = y_{t+h|t,i} - y_{t+h}$  and in which  $y_{t+h|t,i}$  is the out-of-sample forecast of model  $i$ .

For in-sample comparisons, we consider the Deviance Information Criteria (DIC) of Spiegelhalter et al. (2002) as a goodness-of-fit indicator. Generalizations of DIC to missing data problems and to mixture models have been analyzed in Celeux et al. (2006). The DIC is defined as a combination of a classical estimate of fit, based on the likelihood function and a penalty term,  $p_D$ , that measures the complexity of the model. In our application  $p_D = E_{M,S,\theta|Y} [-2 \log p(\mathbf{Y}|M, \mathbf{S}, \theta)] + 2 \log [p(\mathbf{Y}|\hat{M}, \hat{\mathbf{S}}, \hat{\theta})]$  in which  $\hat{\theta}$ ,  $\hat{M}$  and  $\hat{\mathbf{S}}$  can be the estimated posterior mean, median or mode and then

$$\text{DIC} = -2 \log p(\mathbf{Y}|\hat{M}, \hat{\theta}, \hat{\mathbf{S}}, \mathbf{X}) + p_D.$$

In particular,  $(\hat{M}, \hat{\theta})$  are the posterior means obtained from the MCMC recursions, whereas  $\hat{\mathbf{S}}$  is the posterior mode of  $p(\mathbf{S}|\mathbf{Y})$ . The expected likelihood  $E_{M,S,\theta|Y} [-2 \log p(\mathbf{Y}|M, \theta, \mathbf{S}, \mathbf{X})]$  can be estimated by averaging with respect to the MCMC output. It is worth noting that the best model is the one with the smaller DIC.

## 6. Simulation results

In this section, we provide some illustrative examples to show the performance of the algorithm. The goal of this section is threefold. First, we will verify the ability of the algorithm to estimate the parameters and the latent process. Second, we will aim at determining which representation of the ARFIMA dynamics, namely AR and MA, provides a better approximation of the true model. Finally, we would like to estimate the truncation parameter that is fundamental in the determination of the long-range properties of the series. We simulate a time series of length  $T = 1000$  from

$$y_t = \mu_{S_t} + (1 - L)^{-d_{S_t}} \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2). \tag{10}$$

The true parameters, consistent with empirical findings on real data, are  $\mu_0 = 1, \mu_1 = 2, d_0 = 0.35, d_1 = 0.20, \sigma_0^2 = 0.15, \sigma_1^2 = 0.60, p_{10} = p_{01} = 0.01$ . Furthermore, the choice of  $p_{10}$  and  $p_{01}$  guarantees a similar number of contiguous observations in the two regimes. Results, based on 10,000 iterations of the MCMC algorithm with a burn-in of 2500, are reported in Table 2 and in Fig. 2.

Empirical evidence suggests that the algorithm accurately estimates all the parameters of the model. The estimated posterior means are close to the true parameters and are always included in their credibility intervals. In particular, the MA representation seems to help to identify the long range parameters  $d_i$ . This is evident from the last two rows of Fig. 2 in which the estimated posterior distribution of  $d_i$  is displayed. It is clear that the AR approximation provides a poor posterior density for  $d_i$  and especially for  $d_0$ . Regarding the accuracy of the regime estimates, we found that the MA approximation provides a percentage of agreement between the true and the estimated states close to 97%, whereas the AR version 89%. A graphical evidence of this finding is shown in the second row of Fig. 2.

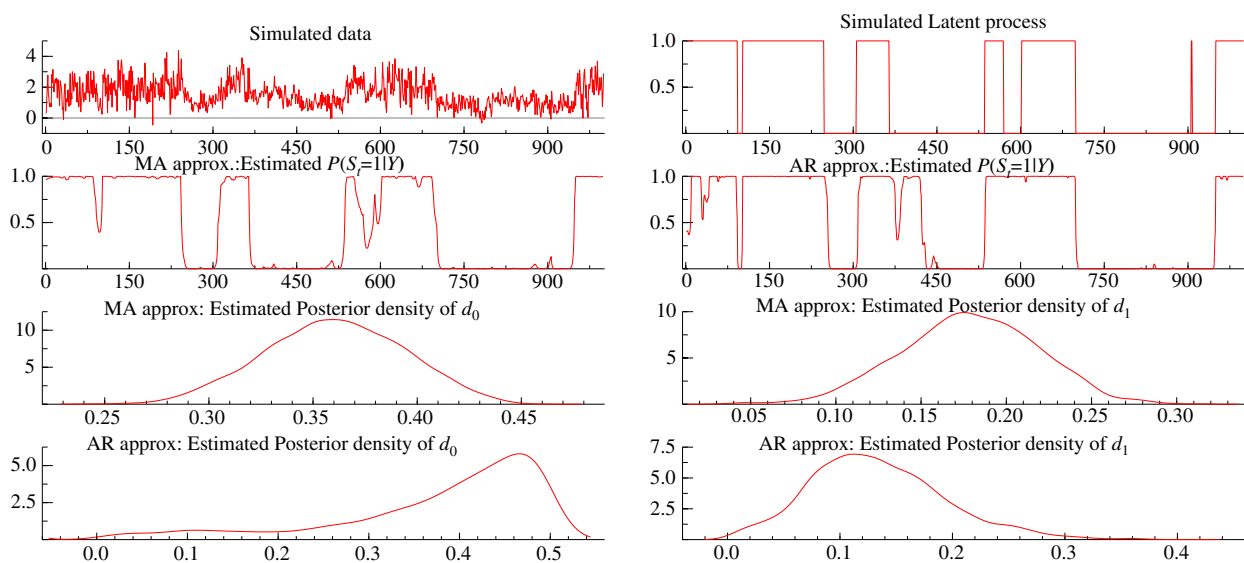
Since estimates are based on the state space representation provided by Eqs. (8)–(9), our third goal is to determine the truncation parameter  $M$ . Chan and Palma (1998) and Ray and Tsay (2002) suggest using an MA(10) approximation for a time



**Table 2**

Parameter estimates of the posterior means. The 95% credible intervals are presented between brackets. log.lik. is obtained by averaging the likelihood with respect to the MCMC draws.

	True param.	MA(M) approx.		AR(M) approx.	
$\mu_0$	1.00	1.0952	[0.918, 1.280]	0.9565	[0.907, 1.010]
$\mu_1$	2.00	1.8923	[1.692, 2.071]	1.9661	[1.901, 2.039]
$d_0$	0.35	0.3602	[0.294, 0.423]	0.3796	[0.060, 0.496]
$d_1$	0.20	0.1782	[0.099, 0.252]	0.1335	[0.028, 0.266]
$\sigma_0^2$	0.15	0.1659	[0.144, 0.191]	0.1858	[0.160, 0.214]
$\sigma_1^2$	0.60	0.5928	[0.519, 0.676]	0.5445	[0.480, 0.614]
$p_{01}$	0.01	0.0188	[0.008, 0.032]	0.0223	[0.011, 0.037]
$p_{10}$	0.01	0.0195	[0.008, 0.034]	0.0175	[0.008, 0.029]
$M$		23.9498	[17.00, 32.00]	17.4020	[10.00, 27.00]
DIC		1845.2		1930.5	
log. lik.		-906.0		-952.0	



**Fig. 2.** Simulated data: Realized volatility and estimated posterior probability of  $\Pr(S_t = 1|Y)$ . On the upper panels are reported the observed realized volatility and the true latent process  $S_t$ . On the second row we report  $\Pr(S_t = 1|Y)$  computed respectively through the MA and the AR approximation. In the last two rows we report the estimated posterior densities of  $d_0$  and  $d_1$ .

series of length 1000. We estimate a slightly higher truncation level  $M$ , thus suggesting that a reasonable tradeoff between accuracy and computational burden can be based on a truncation close to  $M = 20$ , since it belongs to both the credibility intervals. These results are corroborated by a number of Monte Carlo experiments, not reported here but available upon request. It is worth noting that in the case of a few contiguous observations in the same regime, it is difficult to identify the parameters  $\mu_i$  and  $\sigma_i^2$ . Suppose, for instance, there are a few observations on regime 1, then we tend to underestimate the intercept  $\mu_1$  and we overestimate  $\sigma_1^2$ . Further results, not reported here, suggest that this lack of identification is likely due to the small number of contiguous observations labeled by 1.

### 7. Empirical application: S&P500 realized volatility

The empirical analysis is based on 5 min returns of the Standard & Poor's 500 index (S&P500) observed from 1 January 2000 to 28 February 2005. Our MCMC procedure has been written using the Ox@5.0 language of Doornik (2001), combined with the state space library *ssfpack* of Koopman et al. (1999). We ran the algorithm for 10,000 iterations with a burn-in of 2500. In our experience, that is an adequate choice to remove the dependence on the initial conditions. Results are reported in Table 3. An estimate of the switching regime process for the different models is given in Fig. 3.

From Table 3 we notice that all the models with VIX outperform the others, in terms of smaller DIC and higher average likelihood, thus stressing the relevance of the implied volatility as a predictor. A careful look at Table 3 also evidences that parameters of Model 1 characterizing conditional means and variances, i.e.  $\mu_i$  and  $\sigma_i^2$ , are sensibly different across regimes;  $\beta_i$  is quite stable whereas  $d_1$  is slightly bigger than  $d_0$ . These results are consistent with He and Maheu (2010) who find a partial structural break on the NASDAQ volatilities specification in which only the intercept of the conditional variance equation has breaks. As a consequence, Model 1 and Model 1A are substantially equivalent from a goodness-of-fit perspective. It is interesting to note that the non-switching long memory model with VIX, namely Model 1B, provides the better fit in terms of DIC, even though the average likelihood of its switching counterparts is larger. This result can be likely explained by a

**Table 3**

Parameter estimates of the posterior means. The 95% credible intervals are presented between brackets. log.lik. is obtained by averaging the likelihood with respect to the MCMC draws.

	Model 1	Model 1A	Model 1B	Model 1C	Model 2	Model 2A	Model 2B	Model 3
$\mu_0$	−0.1669 [−0.362, 0.002]	−0.1202 [−0.333, 0.073]	−0.4167 [−0.631, −0.208]	0.0619 [−0.011, 0.135]	1.2837 [1.134, 1.434]	1.2916 [1.158, 1.423]	1.5354 [1.356, 1.713]	1.1891 [1.162, 1.218]
$\mu_1$	0.2917 [−0.317, 0.821]	0.2690 [−0.030, 0.540]		−0.0885 [−0.286, 0.249]	1.7607 [1.513, 1.990]	1.7386 [1.528, 1.975]		2.2711 [2.189, 2.349]
$\beta_0$	0.8888 [0.787, 0.998]	0.8612 [0.755, 0.979]	1.0788 [0.973, 1.188]	0.7245 [0.680, 0.770]				
$\beta_1$	0.8366 [0.613, 1.092]			1.0951 [0.945, 1.188]				
$d_0$	0.2057 [0.147, 0.260]	0.2310 [0.178, 0.288]	0.2742 [0.241, 0.308]		0.3917 [0.346, 0.439]	0.3736 [0.339, 0.406]	0.4068 [0.379, 0.435]	
$d_1$	0.3004 [0.190, 0.381]				0.3507 [0.278, 0.428]			
$\sigma_0^2$	0.0915 [0.079, 0.106]	0.0906 [0.077, 0.108]	0.1910 [0.176, 0.206]	0.0866 [0.074, 0.100]	0.0904 [0.077, 0.105]	0.0922 [0.079, 0.107]	0.2129 [0.196, 0.230]	0.1290 [0.115, 0.143]
$\sigma_1^2$	0.5039 [0.410, 0.620]	0.5095 [0.413, 0.624]		0.4464 [0.376, 0.532]	0.5041 [0.419, 0.613]	0.5116 [0.423, 0.622]		0.6038 [0.527, 0.690]
$p_{01}$	0.0261 [0.014, 0.040]	0.0265 [0.015, 0.040]		0.0292 [0.017, 0.044]	0.0291 [0.016, 0.045]	0.0285 [0.016, 0.044]		0.0161 [0.008, 0.026]
$p_{10}$	0.0732 [0.040, 0.116]	0.0725 [0.040, 0.113]		0.0691 [0.0410, 0.102]	0.0645 [0.035, 0.104]	0.0661 [0.035, 0.106]		0.0280 [0.015, 0.044]
$M$	27.4872 [25.0, 32.0]	27.6209 [25.0, 33.0]	28.4075 [26.0, 37.0]		36.9239 [31.0, 39.0]	36.9755 [32.0, 40.0]	36.8837 [34.0, 40.0]	
DIC	1507.4	1534.5	1490.6	1559.2	1749.8	1715.9	1699.0	1908.4
log. lik	−703.1	−707.6	−744.1	−713.2	−797.0	−797.1	−847.2	−932.6

high penalty factor  $p_D$  that accounts for the superior complexity of regime switching models, in which a latent process has to be estimated. However, on closer inspection, we observe sensibly different estimates for *Model 1B* with respect to the non-constrained versions, which suggests some misspecification in *Model 1B*. Regarding the long memory issue, we notice that for models in class 1, the estimates of  $d_i$  are quite moderate compared with the descriptive results of *Table 1*. This finding is consistent with *Bandi and Perron (2006)* in which they provide evidence of a fractionally cointegrated relation between RV and VIX. This result is also confirmed by the results on models of class 2, by observing that the persistence parameters almost double, at least when volatility is low, moving from about 0.20 to 0.40. The higher level of persistence  $d$  is observed for *Model 2B*, in which the regime shifts are not considered, indicating that the long memory coefficients fit the nonlinear effects characterizing the data. As evidenced for models in class 1, also for models in class 2, it appears that the non-switching models provide a superior fit. However, a more accurate look at *Table 3* suggests a clear moving of the conditional variance parameters and a sensitive level shift explained by the Markov switching dynamics, thus suggesting that the more appropriate models are *Model 1* and *Model 2A*. *Model 3* provides evidence that the exclusion of long memory can be also troublesome, since the model provides a poor fit. In particular, our results suggest the utility for joint modeling nonlinearities and high persistence.

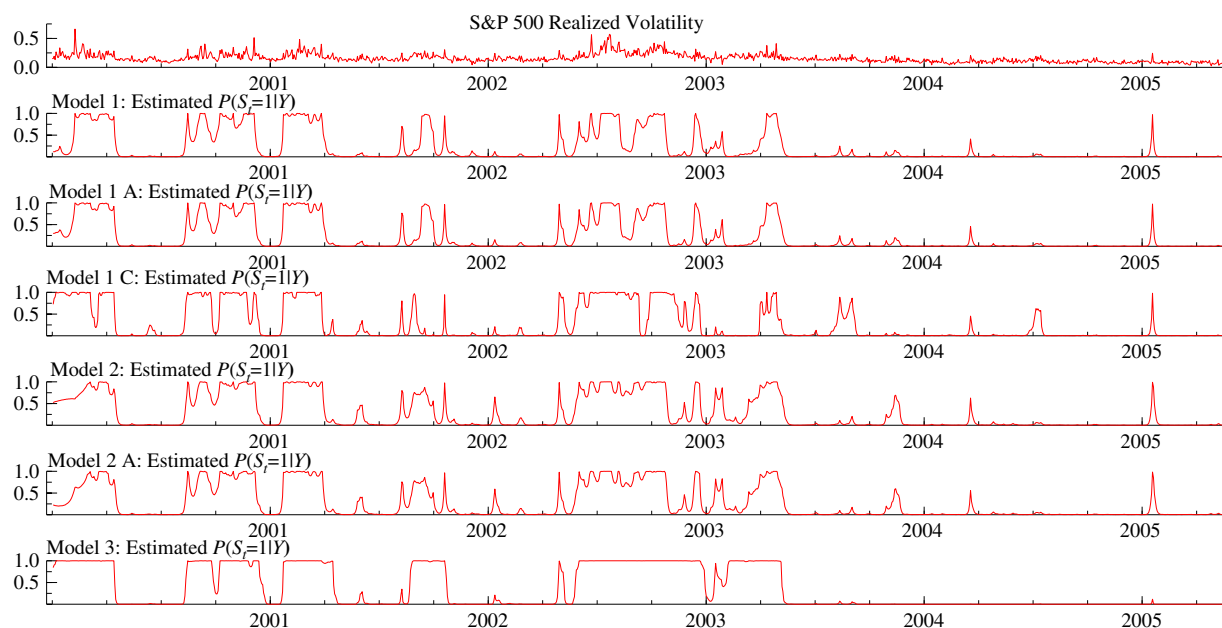
*Fig. 3* shows that all the models considered provide similar regimes estimates, by detecting systematically a regime of high volatility at the beginning and at the end of 2000, at the beginning of 2001 and during mid 2002. Apart for the pure switching model, all the others detect an occasional spike dated 3 November 2004. These results are in line with a graphical inspection of the original data set.

We will now analyze the out-of-sample performance of the models considered. Predictive densities at different horizons have been computed through a sequence of estimates based on a subsample of fixed dimension  $N$ . In our study, the full sample consists of  $T = 1241$  observations, the window is of length  $N = 950$  and the forecasting horizons are  $h = 1, 5, 20$ , corresponding to one trading day, one week and a trading month, leading to a sequence of, respectively,  $K_h = 291, 287, 272$  prediction densities. To avoid the problem of forecasting *VIX*, density forecasts are obtained from

$$y_t = \mu_{S_t} + \beta_{S_t} VIX_{t-h} + (1 - L)^{-d_{S_t}} \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{S_t}^2) \tag{11}$$

and its special cases. A similar forecasting procedure has been considered in *Corsi and Renò (2010)*. First, we report in *Table 4* some classic forecasting accuracy indicators, namely RMSE and MAE. In this Table, we also consider a linear ARMA model, for which the order has been selected following a standard Box–Jenkins procedure.

From *Table 4*, it is clear that the pure switching model provides a poor forecasting performance. Furthermore, the ARMA specification seems to give bad forecasts at a short horizon, whereas it is competitive with the other specifications in the long range, at least with respect to models in class 2. This peculiar behavior is likely induced by the different procedures we followed to compute the ARMA forecasts. In fact, we estimated this ARMA specification through standard maximum likelihood, and we computed the point forecasts by minimizing the RMSE as usual. On the other side, predictive densities for the other models have been evaluated numerically with a procedure affected by Monte Carlo errors. In general, there is



**Fig. 3.** S&P500: realized volatility and estimated posterior probability of  $\Pr(S_t = 1|Y)$ . On the upper panel the *observed* realized volatility. From second to sixth row we report  $\Pr(S_t = 1|Y)$  for all the switching models introduced in Section 3.

**Table 4**  
RMSE and MAE: 1 to 20 days-ahead forecasts.

	1 day-ahead		5 days-ahead		20 days-ahead	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Model 1	0.25738	0.19129	0.30949	0.25294	0.38205	0.31509
Model 1A	0.25742	0.19191	0.31305	0.25623	0.39064	0.32463
Model 1B	0.25535	0.18728	0.29852	0.24018	0.38792	0.31821
Model 1C	0.26183	0.20066	0.31141	0.25603	0.38842	0.31912
Model 2	0.30060	0.23738	0.38958	0.32994	0.56087	0.50426
Model 2A	0.30143	0.23852	0.39244	0.33273	0.56522	0.50910
Model 2B	0.30387	0.24084	0.40336	0.34496	0.62210	0.56797
Model 3	0.43728	0.38111	0.48783	0.43347	0.62269	0.57261
ARMA(2, 1)	0.36182	0.26693	0.36565	0.28004	0.43394	0.34482

a clear evidence that models of class 1 provide superior forecasting performance even at longer horizons. In this case, VIX seems to be more important than the Markov switching dynamics. In fact, all the models of Class 1 outperform the others for each forecasting horizons. Furthermore, even if all models of class 1 provide similar RMSE and MAE for each  $h$ , we observe a slight dominance of *Model 1B*. These findings suggest that the introduction of a regressor is crucial, in order to obtain good results. However, the long memory dynamics improve the forecasting ability of the model. This is clear from [Table 4](#) by observing *Model 1B*. According to the RMSE criteria, the long memory with VIX is superior with respect to *Model 1* and *Model 1C*, for  $h = 1, 5$ . In the long range, i.e., for  $h = 20$ , *Model 1* is superior, meaning that fractional integration together with changes in regimes provides an improvement. The ordering for models in class 2 is slightly different. In fact, *Model 2* is superior to the others at each forecasting horizons. This means that the contribution of the long memory and the switching effects are important. The impact of the switching process is also relevant by observing that the quality of the performance of *Model 2B* decays sensibly when  $h$  increase with respect to the other models. Comparing models of class 1 with models of class 2 through RMSE and MAE, we thus notice that VIX absorbs the switching effects, making its contribution irrelevant for forecasting purposes. However, when VIX is not included, there is strong evidence in favor of long memory together with changes in regimes.

We also provide a pairwise comparisons between the models to shed some light on the importance, in term of forecasting, of the various factors characterizing the volatility dynamics, that is, presence of regressors, nonlinearities and long memory. In fact, preliminary results on forecasting based on standard accuracy statistics are partly supported by a more detailed analysis based on the WLR test of [Amisano and Giacomini \(2007\)](#). Results are included in [Tables 5–7](#) in which pairwise comparisons and associated  $p$ -values in parentheses are reported. A negative value of the test indicate that the alternative model, reported in the columns of the tables, provide a superior predictive accuracy with respect to the benchmark model reported in the rows.

A deeper analysis on WLR evidences that VIX is a good predictor of the S&P500 realized volatility. Evidence reported for different forecasting horizons suggest that models with VIX are, in general, superior to their non-VIX counterparts. In

**Table 5**

Density forecast pairwise comparisons: 1-day-ahead forecasts: (Amisano and Giacomini, 2007) tests and p-values (in parentheses).

Benchmark	Alternative model					
	Mod 1A	Mod 1B	Mod 1C	Mod 2	Mod 2A	Mod 2B
Mod 1	−5.511 (<0.000)	9.415 (<0.000)	−0.089 (0.928)	16.764 (<0.000)	94.853 (<0.000)	12.029 (<0.000)
Mod 1A	– (–)	10.023 (<0.000)	0.159 (0.873)	22.162 (<0.000)	558.76 (<0.000)	12.638 (<0.000)
Mod 1B		– (–)	−85.876 (<0.000)	−7.571 (<0.000)	−6.488 (<0.000)	27.885 (<0.000)
Mod 1C			– (–)	4.997 (<0.000)	3.761 (0.000)	51.166 (<0.000)
Mod 2				– (–)	−1.322 (0.185)	11.044 (<0.000)
Mod 2A					– (–)	9.6052 (<0.000)

**Table 6**

Density forecast pairwise comparisons: 5-day-ahead forecasts: (Amisano and Giacomini, 2007) tests and p-values (in parentheses).

Benchmark	Alternative model					
	Mod 1A	Mod 1B	Mod 1C	Mod 2	Mod 2A	Mod 2B
Mod 1	0.552 (0.580)	16.801 (<0.000)	−1.405 (0.159)	7.474 (<0.000)	6.955 (<0.000)	15.541 (<0.000)
Mod 1A	– (–)	18.987 (–)	−1.851 (0.064)	8.218 (<0.000)	7.651 (<0.000)	16.510 (<0.000)
Mod 1B		– (–)	−48.919 (<0.000)	−34.750 (<0.000)	−39.959 (<0.000)	14.219 (<0.000)
Mod 1C			– (–)	17.200 (<0.000)	16.290 (<0.000)	23.765 (<0.000)
Mod 2				– (–)	−64.359 (<0.000)	28.989 (<0.000)
Mod 2A					– (–)	29.513 (<0.000)

**Table 7**

Density forecast pairwise comparisons: 20-day-ahead forecasts: (Amisano and Giacomini, 2007) tests and p-values (in parentheses).

Benchmark	Alternative model					
	Mod 1A	Mod 1B	Mod 1C	Mod 2	Mod 2A	Mod 2B
Mod 1	4.316 (<0.000)	22.220 (<0.000)	−0.937 (0.348)	12.886 (<0.000)	13.026 (<0.000)	15.196 (<0.000)
Mod 1A	– (–)	20.630 (<0.000)	−1.119 (0.262)	12.387 (<0.000)	12.517 (<0.000)	14.847 (<0.000)
Mod 1B		– (–)	−8.271 (<0.000)	4.667 (<0.000)	4.732 (<0.000)	12.384 (<0.000)
Mod 1C			– (–)	7.257 (<0.000)	7.293 (<0.000)	10.174 (<0.000)
Mod 2				– (–)	−2.032 (0.042)	18.821 (<0.000)
Mod 2A					– (–)	18.492 (<0.000)

particular, models of class 1 are in general better than models of class 2. This finding is true for each forecasting horizon considered. However, it is interesting to note that *Model 1B* provides a poor forecasting performance, even if compared to some models of class 2, namely *Model 2* and *Model 2A*. This finding evidences that the inclusion of VIX does not always provide a forecasting ability improvement and that switching dynamics may compensate for the regressor omission, at least for  $h = 1$  and 5.

Regarding models of class 1, we notice that *Model 1* and *Model 1C* provide equivalent performance, for  $h = 1, 5$  and 20, which means that when VIX is included, the long memory effects are not relevant. In particular, these two models clearly outperform *Model 1A* and *Model 1B* in the long range. This finding is substantially in line with former results based on RMSE, i.e., the VIX effect absorbs long range dynamics due to its high level of persistence. A surprising result, that contrasts with former evidence based on RMSE and MAE, shows that the switching dynamics is important to predict future realizations of the realized volatility and in fact, for each  $h$ , *Model 1C* outperforms *Model 1B*. This point is stressed from model from class 2, in which we find that *Model 2* and *Model 2A* outperforms *Model 2B*. These results indicate that when VIX is not included, the switching dynamics, together with the long memory, have to be included in the model.

According to these results, it is clear that the most important factor in providing reasonable forecasts is the presence of predictors in the model. However, it seems that the combination of long memory, together with switching regimes, represents an improvement with respect to the pure switching and the pure long memory dynamics.

## 8. Conclusions

In this paper we have modeled realized volatility through regime switching models and long range persistence. To perform inference we proposed an MCMC algorithm based on the MA representation of the ARFIMA dynamics. We also included in our specification exogenous regressors, i.e., the implied volatility, to improve the forecasting performance of the models proposed. Analyses based on simulated and real data, evidence that realized volatility is characterized by long memory and changes in regimes. Furthermore, implied volatility dramatically improves the in-sample and the out-of-sample results. However, the Amisano and Giacomini (2007) test evidences that the forecasting performance is also affected by the introduction of the long memory and of the Markov switching dynamics.

## Acknowledgements

Financial support from the MIUR under grant PRIN 2008 Prot. N. 2008MRFM2H\_003 is gratefully acknowledged. We wish to thank two anonymous referees and the Associate Editor for helpful comments, David Ardia, Roberto Casarin, the participants at the *Second International Workshop on Computational and Financial Econometrics* held in Neuchâtel on June 2008 and the participants at the *14th International Conference on Computing in Economics and Finance* held in Paris on June 2008. All errors are our own.

## References

- Albert, J.H., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics* 11, 1–15.
- Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25, 177–190.
- Andersen, T., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T.G., Bollerslev, T., Diebold, F., Labys, P., 2001. The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T.G., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Ardia, D., 2009. Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with student-*t* innovations. *Econometrics Journal* 12, 105–126.
- Baillie, R., Kapetanios, G., 2007. Testing for neglected nonlinearity in long memory models. *Journal of Business & Economic Statistics* 25, 447–461.
- Bandi, F.M., Perron, B., 2006. Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics* 4, 636–670.
- Bandi, F.M., Russell, J., 2008. Microstructure noise, realized volatility, and optimal sampling. *Review of Economic Studies* 75, 339–369.
- Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B* 64, 253–280.
- Billio, M., Monfort, A., Robert, C., 1999. Bayesian estimation of switching ARMA models. *Journal of Econometrics* 93, 229–255.
- Brockwell, A., 2006. Likelihood based analysis of a class of generalized long memory models. *Journal of Time Series Analysis* 28, 386–407.
- Cai, J., 1994. A Markov model of unconditional variance in ARCH. *Journal of Business & Economic Statistics* 12, 309–316.
- Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Carvalho, C., Lopes, H., 2007. Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis* 51, 4526–4542.
- Celeux, G., Forbes, F., Robert, C., Titterton, D., 2006. Deviance information criteria for missing data models. *Bayesian Analysis* 1, 651–674.
- Chan, N.H., Palma, W., 1998. State space modeling of long-memory processes. *Annals of Statistics* 26, 719–740.
- Chen, C., Gerlach, R., Lin, E.M., 2008. Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics & Data Analysis* 52, 2990–3010.
- Chen, C., Yu, T., 2005. Long-term dependence with asymmetric conditional heteroscedasticity in stock returns. *Physica A* 353, 413–424.
- Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* 75, 79–97.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Corsi, F., Renò, R., 2010. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Working Paper. Università di Siena, Dipartimento di Economia Politica.
- Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131–159.
- Doornik, J., 2001. *Ox: An Object-Oriented Matrix Programming Language*. Timberlake Consultants Press, London.
- Doornik, J.A., Ooms, M., 2006. A package for estimating, forecasting and simulating ARFIMA models: ARFIMA package 1.04 for Ox. Tech. Rep. Nuffield College.
- Dueker, M., 1997. Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business & Economic Statistics* 15, 26–34.
- Ehlers, R., Brooks, S., 2004. Bayesian analysis of order uncertainty in ARIMA models. Technical Report. University of Cambridge.
- Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 96, 194–209.
- Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. Springer, Berlin.
- Granger, C., Ding, Z., 1996. Varieties of long memory models. *Journal of Econometrics* 73, 61–77.
- Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
- Haas, M., Mittnik, S., Paolella, M., 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2, 493–530.
- Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
- Hamilton, J., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.
- He, Z., Maheu, J.M., 2010. Real time detection of structural breaks in GARCH models. *Computational Statistics & Data Analysis* 54, 2628–2640.
- Henneke, J., Rachev, S., Fabozzi, F., Nikolov, M., 2009. MCMC-based estimation of Markov switching ARMA–GARCH models. *Applied Economics* iFirst 1–13.
- Hillebrand, E., Medeiros, M., 2008. Asymmetries, breaks, and long-range dependence: an estimation framework for time series of daily realized volatility. Discussion Paper. Pontifical Catholic University of Rio de Janeiro.
- Hsu, N., Breidt, F., 2003. Bayesian analysis of fractionally integrated ARMA with additive noise. *Journal of Forecasting* 22, 491–514.
- Kauffmann, F., Frühwirth-Schnatter, S., 2002. Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis* 23, 425–458.



- Kim, C.J., Nelson, C., 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press.
- Klaassen, F., 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27, 363–394.
- Ko, K., Vannucci, M., 2006. Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference* 136, 3415–3434.
- Koop, G., Ley, E., Osiewalski, J., Steel, M.F.J., 1997. Bayesian analysis of long memory and persistence using ARFIMA models. *Journal of Econometrics* 76, 149–169.
- Koopman, S., Jungnaker, B., Hol, E., 2005. Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.
- Koopman, S., Shephard, N., Doornik, J.A., 1999. Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* 2, 113–166.
- Longin, F., 1997. The threshold effect in expected volatility: a model based on asymmetric information. *Review of Financial Studies* 10, 837–869.
- Lux, S., Morales-Arias, L., 2010. Forecasting volatility under fractality, regime-switching, long memory and Student-*t* innovations. *Computational Statistics & Data Analysis* 54, 2676–2692.
- Marcucci, J., 2005. Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 9, 1–53.
- Martens, M., van Dick, D., de Pooter, M., 2004. Modeling and forecasting S&P500 volatility: long memory, structural breaks and nonlinearity. Discussion Paper. ti 2004-067/4, Tinbergen Institute.
- McAleer, M., Medeiros, M., 2008a. A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics* 147, 104–119.
- McAleer, M., Medeiros, M., 2008b. Realized volatility: a review. *Econometric Reviews* 27, 10–45.
- McCulloch, R., Tsay, R.S., 1993. Statistical inference of macroeconomic time series via Markov switching models. *Journal of Time Series Analysis* 15, 523–539.
- Newey, W., West, K., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Ohanissian, A., Russell, J., Tsay, R., 2008. True or spurious long memory? A new test. *Journal of Business & Economic Statistics* 26, 161–175.
- Pai, J.S., Ravishanker, N., 1996. Bayesian modelling of ARFIMA processes by Markov chain Monte Carlo methods. *Journal of Forecasting* 15, 63–82.
- Pai, J.S., Ravishanker, N., 1998. Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis* 19, 99–112.
- Palma, W., 2007. *Long-Memory Time Series: Theory and Methods*. In: *Wiley Series in Probability and Statistics*, J. Wiley & Sons Ed., Hoboken, NJ.
- Perron, B., Qu, Z., 2010. Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics* 28, 275–290.
- Ray, B.K., Tsay, R.S., 2002. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* 23, 687–705.
- Robert, C., Casella, G., 1999. *Monte Carlo Statistical Methods*. Springer, Berlin.
- Shephard, N., 1994. Partial non-Gaussian state space. *Biometrika* 81, 115–131.
- Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with comments). *Journal of the Royal Statistical Society. Series B* 64, 583–639.
- Susmel, R., 2000. Switching volatility in private international equity markets. *International Journal of Finance & Economics* 5, 265–283.
- Troughton, P., Godsill, S., 1998. A reversible jump sampler for autoregressive time series. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*. vol. 4. pp. 2257–2260.
- Vermaak, J., Andreau, C., Doucet, A., Godsill, S., 2004. Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis* 25, 785–809.