UNIVERSITA' DEGLI STUDI DI BOLOGNA

DIPARTIMENTO DI SCIENZE ECONOMICHE

Long memory and nonlinearities in realized volatility: a Markov switching approach

Silvano Bordignon
(University of Padova)

Davide Raggi

Quaderni - Working Papers DSE N° 694

Long memory and nonlinearities in realized volatility: a Markov switching approach

Silvano Bordignon *
University of Padova

Davide Raggi†
University of Bologna

February 6, 2010

Abstract

Goal of this paper is to analyze and forecast realized volatility through nonlinear and highly persistent dynamics. In particular, we propose a model that simultaneously captures long memory and nonlinearities in which level and persistence shift through a Markov switching dynamics. We consider an efficient Markov chain Monte Carlo (MCMC) algorithm 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 from nested models.

Keywords: Realized volatility, Switching-regime, Long memory, MCMC, Forecasting

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 compared to other volatility measures such as squared or absolute daily returns. Thus, volatility becomes in some sense *observable* and can be modeled

*University of Padova, Department of Statistics, Via C. Battisti 33, 35123 Padova, Italy. E-mail: `bordigno@stat.unipd.it`.

†University of Bologna, Department of Economics, Piazza Scaravilli 2, 40126 Bologna, Italy. E-mail: `davide.raggi@unibo.it`.

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 the recent literature and a detailed review has been provided in McAleer and Medeiros (2008b) for instance. One of the most relevant is that RV dynamics exhibits long memory or high persistence, as evidenced, amongst others, in Andersen et al. (2003), Corsi (2009) and Koopman et al. (2005). 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. However, Granger and Ding (1996) found out that persistence in volatility tends to be non constant over time and in particular Longin (1997) provided evidence of an usually higher level of persistence when volatility is low, thus suggesting presence of nonlinearities.

On the other hand, it is well known that long memory can be overestimated when 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 aiming at disentangle the effects of long memory and level shifts has been proposed in Baillie and Kapetanios (2007) and in Ohanissian et al. (2008). In particular, Baillie and Kapetanios (2007) found presence of nonlinearity together with long memory in realized volatilities for currencies, whereas Ohanissian et al. (2008), in their empirical application, provide evidence that realized volatility of exchange rates such as DM/$ and Yen/$ is properly described by a true long memory process. This latter result seems to be in contrast with Perron and Qu (2009), 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 (2009) propose a GARCH model subject to an unknown number of structural breaks.

It seems to be still an open question whether to decide if long memory is spurious or not. It thus appears of obvious interest to joint modeling both 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 been suggested to combine long memory and nonlinearities to describe conditional variances. Probably, the first contributions in this direction are, respectively, Martens et al. (2004) and Hillebrand and Medeiros (2008) where they build from an ARFIMA model by allowing for smooth level shifts, day of the week effects and leverage. A different strategy has been proposed in McAleer and Medeiros (2008a) by introducing a multiple regime smooth transition extension of the Heterogeneous Autoregressive (HAR) model of Corsi (2009).

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 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 also 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.

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

We base our empirical analysis on the 5 minutes intra-daily series of Standard & Poor's 500 (S&P500) stock index over the period January 2000 to 28 February 2005. Our results evidence that implied volatility is important to predict RV and also that, in the short run, long memory together with nonlinearities improve the 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, whereas the Markov switching models considered are defined in Section 3. 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 respectively in Section 6 and in Section 7.

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 deeply 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

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

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 Integrated Volatility at day t, namely, IV_t.

It can be proved that, in absence of *microstructure noise*, i.e., the observed prices are not affected by measurement errors, IV can be consistently estimated by the realized volatility, defined as

$$\hat{y}_t = RV_t = \sum_{j=1}^{N_t} \left[p_{j,t} - p_{j-1,t} \right]^2, \qquad j = 1, \ldots, N_t, \ t = 1, \ldots, T \tag{1}$$

in which $p_{j,t}$ is the j-th observation at day t and N_t is the number of intra-daily observations.

In case of microstructure noise, the true prices $p^*_{t,j}$ are affected by measurement errors, and then $p_{j,t} = p^*_{j,t} + \epsilon_{j,t}$. As a main consequence, the realized volatility estimator diverges, in case 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 become negligible and then a sampling frequency ranging between 5 to 30 minutes can be a reasonable choice. See also Brownlees and Gallo (2006) for a more accurate treatment on data handling with ultra-high frequency data.

We also consider in our study the CBOE Volatility Index (VIX) that 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 is referred to option contracts, it can be seen as a market's predictor of the expected volatility.

The S&P500 index data set consists on 5 minutes intra-daily observations whereas the VIX index is observed on a daily basis from 1 January 2000 to 28 February 2005. Prices of S&P500 have been provided by Olsen and Associates in Zurich and realized volatility has been thus computed through formula (2) in which $N_t = 288$. We have removed from our sample days in which market has been closed, such as the three weeks after September 11, week-ends and US holidays, leading to a total amount of $T = 1274$ trading days. The VIX index has been downloaded from the `CBOE` website[1].

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 whereas Figure 1 displays the dynamics of the volatility measures. In particular from Figure 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.

[1]The data set for VIX is available at `http://www.cboe.com/micro/vix/introduction.aspx`

Table 1: Descriptive statistics of RV and VIX. We report also the estimated d of the ARFIMA(0,d,0) model obtained through Maximum Likelihood.

	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	0.499673

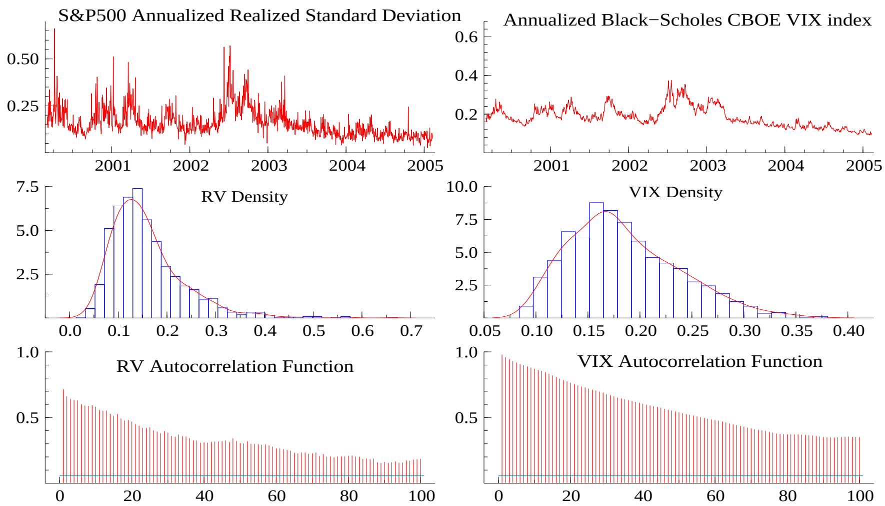

Figure 1: Realized standard deviations and VIX index data from 01/03/2000 to 28/02/2005, estimated densities and autocorrelation function.

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, in which 0 identify periods with *low* *volatility* whereas 1 indicate *high volatility*.

Our specification is similar in spirit to Ray and Tsay (2002), in which they first propose an ARFIMA model with random level shifts but with constant d and, second, an ARFIMA model with time varying persistence levels[2] but constant parameters. Here we consider a Markov regime switching model in line with Hamilton (1989) in which the mixture components are fractionally integrated. In our general parameterisation, we allow, even though we do not strictly require, that all the parameters are regime dependent, thus allowing for level shifts together with time varying persistence. The model is specified as follows

$$y_t \quad = \quad \mu_{S_t} + \beta_{S_t} \mathrm{X}_t + (1-L)^{-d_{S_t}} \epsilon_t \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma^2_{S_t}\right), \tag{2}$$

in which $y_t,\ t = 1, \ldots, T$ are the realized volatilities and $S_t,\ t = 1, \ldots, T$ 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, $X_t,\ t = 1, \ldots, T$ is the implied volatility index VIX. We will label this model by **Model 1**. We also consider the following priors $\mu_0 \sim \mathcal{N}(0, 0.5)$, $\mu_1 \sim \mathcal{N}(1, 0.5)$, $\beta_i \sim \mathcal{N}(0, 0.5)$, $d_0 \sim \mathcal{N}(0.1, 0.01)$ $d_1 \sim \mathcal{N}(0.3, 0.01)$, $\sigma^2_i \sim IG(2.5, 0.75)$ and finally $p_{ij} \sim Beta(5, 95)$, for $i, j = 0, 1$. Long memory parameters are also constrained in the interval $(0, 0.5)$. Furthermore, to avoid label switching, a common issue with mixture models, we impose $\sigma^2_1 > \sigma^2_0$. This parameterisation encompasses many other special cases, and in particular we consider a model with common β and persistence, i.e. $d_0 = d_1 = d$, but regime-dependent levels and conditional variances, namely **Model 1 A**,

$$y_t = \mu_{S_t} + \beta \mathrm{VIX}_t + (1-L)^{-d} \epsilon_t \quad \epsilon_t \sim \mathcal{N}\left(0, \sigma^2_{S_t}\right), \tag{3}$$

and the non switching ARFIMA$(0, d, 0)$ with exogenous regressor VIX, that is **Model 1 B**, defined as

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

[2]In particular Ray and Tsay (2002) suggest a persistence level that shifts according to a Markovian process.

We also estimate the same models described above by constraining $\beta_0 = \beta_1 = \beta = 0$, i.e. we omit the exogenous regressor and we call them respectively **Model 2**, **Model 2 A** and **Model 2 B**. Finally **Model 3** is a pure Markov switching model defined as

$$y_t = \mu_{S_t} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma^2_{S_t}\right) \tag{5}$$

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 $\mathbf{S} = (S_1, \ldots, S_T)$ is not observable and also because there is not a standard method to estimate the long range parameters $d_i,\ i = 0, 1$.

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, there is not a unique estimator available for $d_i,\ i = 0, 1$ and a detailed review on inference for long memory parameters has been provided by Palma (2007).

In this paper we focus on MCMC methods for inference. Our solution is based on the idea of Chan and Palma (1998), that prove 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. Furthermore, to make the computations feasible, the Kalman recursions are based on the truncation of the infinite Moving Average (MA) representation of a long memory process. For instance, the ARFIMA(0,d,0) model has a linear MA representation given by

$$y_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \epsilon_{t-j} \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{6}$$

Chan and Palma (1998) consider the approximation of eq. (6) 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 whole regimes vector $\mathbf{S}$, we can represent the dynamics defined in (2), by the time varying parameter state space system given as follows

$$\underbrace{\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \\ \vdots \\ \epsilon_{t-M+1} \end{bmatrix}}_{\alpha_{t+1\,(M+1)\times 1}} = \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}}_{T_{(M+1)\times(M+1)}} \underbrace{\begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-M} \end{bmatrix}}_{\alpha_t} + \underbrace{\begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{u_{t+1\,(M+1)\times 1}} \tag{7}$$

$$y_t = \mu_{S_t} + \beta_{S_t} X_t + \underbrace{\begin{bmatrix} 1 & \pi_1(d_{S_t}) & \pi_2(d_{S_t}) & \dots & \pi_M(d_{S_t}) \end{bmatrix}}_{Z_{t\ 1\times(M+1)}} \alpha_t. \tag{8}$$

in which $\pi_j(d_{S_t}) = \frac{\Gamma(j+d_{S_t})}{\Gamma(j+1)\Gamma(d_{S_t})}$. Given this state space representation and conditional on $\mathbf{S}$, inference for the parameters can be derived by applying standard results on MCMC for ARMA models. Seminal contributions on this field are Albert and Chib (1993) and McCulloch and Tsay (1993) whereas relevant generalizations have been successively proposed in Billio et al. (1999) and in Kim and Nelson (1999). Furthermore, Ray and Tsay (2002) propose a Gibbs sampler algorithm to take into account long range dependencies together with random level shifts. However, to our knowledge, MCMC methodologies have never been implemented for ARFIMA model with switching regime parameters.

We propose an MCMC algorithm to simulate from the posterior distribution of $(\boldsymbol{\theta}, \mathbf{S})$ where $\boldsymbol{\theta} = (\mu_i, \beta_i, \sigma_i, d_i, p_{01}, p_{10})$, $i = 0, 1$, is the parameter's vector. In the following we will call $\boldsymbol{\theta}_j$ the generic j-th element of $\boldsymbol{\theta}$. Finally $\boldsymbol{Y}$ contains the observable realized volatilities, whereas $\boldsymbol{X}$ holds the exogenous regressor's realization.

The basic idea behind MCMC is to simulate a trajectory of a Markov chain $\{\boldsymbol{\theta}^{(j)}, \mathbf{S}^{(j)}\}_{j=1}^{n}$ from a given starting point $(\boldsymbol{\theta}^{(0)}, \mathbf{S}^{(0)})$, with limiting invariant distribution[3] $p(\mathbf{S}, \boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})$. Once convergence is achieved, the algorithm provides a sample of serially dependent draws for $\boldsymbol{\theta}$ and $\mathbf{S}$, which can be used to perform inference. More precisely, estimates of the latent factors and of the posterior mean of $\boldsymbol{\theta}$ are given by averaging over the realization of the chain, i.e. $\widehat{Pr}(S_t = 1) = n^{-1}\sum_{j=1}^{n} S_t^{(j)}$ and $\hat{\boldsymbol{\theta}} = n^{-1}\sum_{j=1}^{n} \boldsymbol{\theta}^{(j)}$. To account for the serial correlation in the draws, the numerical standard error of the sample posterior mean is estimated using the procedure implemented in Kim et al. (1998).

Moving the whole vector $(\boldsymbol{\theta}, \mathbf{S})$ in block can be difficult, since it is highly multivariate. A possible strategy is to divide it into sub-components and then update them one-at-a-time. As suggested in Shephard (1994) and Carter and Kohn (1994) amongst others, updating the whole latent process $\mathbf{S}$ in 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, we take care of the regime switches by providing an efficient algorithm based on the multi-move Gibbs sampler proposed in Chib (1996) to update the states $\mathbf{S}$, whereas parameters are updated individually. We can summarize the algorithm as follows

- Initialize the chain at $(\boldsymbol{\theta}^{(0)}, \mathbf{S}^{(0)})$
- At step $j = 1, \dots, n$
  - Update $\boldsymbol{\theta}$ one-at-a-time from $p(\boldsymbol{\theta}_i|\mathbf{S}^{(j-1)}, \boldsymbol{\theta}_{-i^-}^{(j)}, \boldsymbol{\theta}_{-i^+}^{(j-1)}, \boldsymbol{Y}, \boldsymbol{X})$ through Metropolis-Hastings, where $\boldsymbol{\theta}_{-i^-}$ are the first (i-1) elements of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{-i^+}$ are the elements form the (i+1)-th to the last.
  - Draw $\mathbf{S}^{(j)}$ in block from $p(\mathbf{S}|\boldsymbol{\theta}^{(j)}, \boldsymbol{Y}, \boldsymbol{X})$, using the method of Chib (1996);

[3] See Robert and Casella (1999) ch. 6-7 for a general treatment on the conditions needed to achieve convergence of MCMC algorithms.

In the next subsections we will describe in detail the algorithm.

4.1 Updating the parameters

Simulating p_{01} and p_{10} is straightforward, since it is easy to show that their conditional posterior distributions are Beta. Updating σ_i^2 $i = 0, 1$, is also easy, since their conditional posterior are Inverse Gamma. However, to identify the two states, we need to impose that $\sigma_0^2 < \sigma_1^2$. This leads to a *constrained* conditional posterior distributions that are truncated, but can be simulated trough the accept-reject algorithm of Philippe (1997). Sampling μ_i, β_i and $d_i,\ i = 0, 1$ is more involved, since their conditional distributions are not known in closed form[4]. 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 $\boldsymbol{\theta}_i$ with unknown full conditional using the following scheme

1. Sample $\boldsymbol{\theta}_i^*$ from a proposal $q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{(j-1)})$.
2. Define $\alpha(\boldsymbol{\theta}_i^{(j-1)}, \boldsymbol{\theta}_i^*) = \min\left(\frac{p(\boldsymbol{\theta}_i^*)p(\boldsymbol{Y}|\boldsymbol{\theta}_i^*, \boldsymbol{X}, \mathbf{S}^{(j)}, \boldsymbol{\theta}_{-i^-}^{(j)}, \boldsymbol{\theta}_{-i^+}^{(j-1)})q(\boldsymbol{\theta}_i^{(j-1)}|\boldsymbol{\theta}_i^*)}{p(\boldsymbol{\theta}_i^{(j-1)})p(\boldsymbol{Y}|\boldsymbol{\theta}_i^{(j-1)}, \boldsymbol{X}, \mathbf{S}^{(j)}, \boldsymbol{\theta}_{-i^-}^{(j)}, \boldsymbol{\theta}_{-i^+}^{(j-1)})q(\boldsymbol{\theta}_i^*|\boldsymbol{\theta}_i^{(j-1)})}, 1\right)$
3. Sample u from $U(0, 1)$.
4. If $u \leq \alpha(\boldsymbol{\theta}_i^{(j-1)}, \boldsymbol{\theta}_i^*)$

   then $\qquad \boldsymbol{\theta}_i^{(j)} = \boldsymbol{\theta}_i^*$

   otherwise $\qquad \boldsymbol{\theta}_i^{(j)} = \boldsymbol{\theta}_i^{(j-1)}$.

Here $p(\boldsymbol{\theta}_i)$ is the prior distribution for the i-th parameter, $p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{X}, \mathbf{S})$ is the likelihood of the model conditioned on $\mathbf{S}$, $\boldsymbol{\theta}_{-i^-}$ and $\boldsymbol{\theta}_{-i^+}$. Evaluating $p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{X}, \mathbf{S})$ is not trivial and is computationally demanding. However, since the model can be written in the state-space form provided by eqs. (7)-(8), we exploit the Kalman filter recursions.

[4]Billio et al. (1999) provide a closed form expression for the full conditional for $\mu_i,\ i = 0, 1$. We use a Metropolis-Hastings approach since it gives equivalent results.

4.2 Updating S

We update $\mathbf{S}$ in block, moving the vector according to the algorithm proposed in Chib (1996). The method is based on the following decomposition of $p(\mathbf{S}|\boldsymbol{Y}, \boldsymbol{\theta})$, that is,

$$p(\mathbf{S}|\boldsymbol{Y}, \boldsymbol{\theta}) = p(S_T|\boldsymbol{Y}, \boldsymbol{\theta}) \dots p(S_t|\boldsymbol{Y}, S_{t+1}, \dots, S_T, \boldsymbol{\theta}) \dots p(S_1|\boldsymbol{Y}, S_2, \dots, S_T, \boldsymbol{\theta}), \tag{9}$$

where, in particular $p(S_t|\boldsymbol{Y}, S_{t+1}, \dots, S_T, \boldsymbol{\theta})$ can be written as

$$\begin{aligned} p(S_t|\boldsymbol{Y}, S_{t+1}, \dots, S_T, \boldsymbol{\theta}) \quad &\propto \quad p(S_t|y_1, \dots, y_t, \boldsymbol{\theta}) p(S_{t+1}|S_t, \boldsymbol{\theta}) \\ &\quad\ \ p(S_t|y_1, \dots, y_{t-1}, \boldsymbol{\theta}) p(y_t|y_1, \dots, y_{t-1}, \boldsymbol{\theta}) p(S_{t+1}|S_t, \boldsymbol{\theta}). \end{aligned}$$

The distribution $p(S_t|y_1, \dots, y_{t-1})$ can be decomposed as

$$p(S_t|y_1, \dots, y_{t-1}) = \sum_{j=0}^{1} p(S_t|S_{t-1} = j, \boldsymbol{\theta}) p(S_{t-1} = j|y_1, \dots, y_{t-1}, \boldsymbol{\theta})$$

whereas $p(S_t|y_1, \dots, y_t, \boldsymbol{\theta})$ is given by

$$p(S_t|y_1, \dots, y_t, \boldsymbol{\theta}) \propto p(S_t|y_1, \dots, y_{t-1}) p(y_t|y_1, \dots, y_{t-1}, \boldsymbol{\theta}).$$

Numerical evaluation of $p(y_t|y_1, \dots, y_{t-1}, \boldsymbol{\theta})$ is provided by applying the Kalman filter to the state space approximation described by equations (7) and (8) whereas $p(S_t|y_1, \dots, y_{t-1})$ and $p(S_t|y_1, \dots, y_t, \boldsymbol{\theta})$ can be computed recursively by setting $p(S_1|y_0, \boldsymbol{\theta}) = p(S_1|\boldsymbol{\theta})$, i.e., the stationary distribution of the Markov chain. Once computed all these quantities it is possible to simulate S_T from $p(S_T|\boldsymbol{Y}, \boldsymbol{\theta})$ that is a Binomial random variable, and finally, all the remaining states can be directly simulated from $p(S_t|\boldsymbol{Y}, S_{t+1}, \dots, S_T, \boldsymbol{\theta})$, starting from S_{T-1} until S_1.

4.3 Forecasting

MCMC allows to easily approximate the density forecasts of $y_{T+1}, \dots, y_{T+h}$ for a given horizon $h > 0$. At each Gibbs iteration, we can simulate a 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 $\boldsymbol{\theta}$. To take into account uncertainties about future realizations of the exogenous regressors, we first suppose, for instance, that it follows an $ARFIMA(0, d_x, 0)$ process, that is

$$X_t = \mu_x + (1-L)^{-d_x}\epsilon_t \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_x^2\right).$$

To simplify the notation we indicate future realizations of $\boldsymbol{X}$, $\mathbf{S}$ and $\boldsymbol{Y}$ as $\boldsymbol{X}_h = (X_{T+1}, \ldots, X_{T+h})$, $\mathbf{S}_h = (S_{T+1}, \ldots, S_{T+h})$ and $\boldsymbol{Y}_h = (y_{T+1}, \ldots, y_{T+h})$ respectively whereas $\boldsymbol{\theta}_X = (\mu_x, d_x, \sigma_x^2)$.

Simulating from $p(y_{T+i}|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}),\ i = 1, \ldots, h$ can be done as follows:

- At the j-th Gibbs iteration set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}$;
  - Simulate $\boldsymbol{\theta}_X^{(j)}$ from $p(\boldsymbol{\theta}_X|\boldsymbol{X})$;
  - For $i = 1, \ldots, h$
    * Draw $S_{T+i}^{(j)}$ from $p(S_{T+i}|S_{T+i-1}^{(j)}, \boldsymbol{\theta}^{(j)})$;
    * Draw $X_{T+i}^{(j)}$ from $p(X_{T+i}|\boldsymbol{X}, \ldots, \boldsymbol{X}_{T+i-1}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}_X^{(j)})$;
    * Draw $y_{T+i}^{(j)}$ from $p(y_{T+i}|\boldsymbol{Y}, \boldsymbol{Y}_{T+i-1}^{(j)}, \boldsymbol{X}, \boldsymbol{X}_{T+i}^{(j)}, \mathbf{S}, S_{T+i}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}_X^{(j)})$;
  - End for.
- Record $(y_{T+i}^{(j)})_{j=1}^h$.

The sequences $y_{T+i}^{(j)},\ j = 1, \ldots, n,\ i = 1, \ldots, h$ can be used to estimate the predictive densities $p(y_{t+i}|\boldsymbol{Y})$.

5 Forecasting methodology and in-sample goodness-of-fit

To compare the out-of-sample predictive accuracy of the models, we use the *Weighted Likelihood Ratio test (WLR)* proposed in Amisano and Giacomini (2007) that compares density forecasts computed through classic or Bayesian techniques. The key requirement for the WLR test is that all the forecasts are based on a finite estimation window, that is, the recursion of the h-steps-ahead, $h \geq 1$, density forecasts are obtained using a fixed sample of size N. In practice, the first h-steps-ahead forecast is based on $(y_1, \ldots, y_N)$, the second on $(y_2, \ldots, y_{N+1})$, until the last one, that is based on $(y_{T-N-h+1}, \ldots, y_{T-h})$. We thus obtain a sequence of $K = T - N - h - 1$ 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 $(y_{t-N+1}, \ldots, y_t)$ under models $\mathcal{M}_i$ and $\mathcal{M}_j$ and finally $\omega(\cdot)$ is a weight that allows to emphasize some specific region of the support of the density forecasts[5]. 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

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

and the associated statistic is

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

where $\widehat{\text{avar}}(\overline{WLR}_{t+h})$ 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 $t_h \overset{a}{\sim} \mathcal{N}(0, 1)$.

We also consider the Root Means Square Error (RMSE) together with the Mean Absolute Error (MAE), that is, for model i $\text{RMSE}_i = \sqrt{\frac{1}{K_h}\sum_{t=N}^{T-h} \epsilon^2_{t+h|t,\ i}}$

[5]For example if we are more interested on the center of the distribution $\omega(y) = \phi(y)$, where ϕ is the standard normal density function.

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 as goodness-of-fit statistic the Deviance Information Criteria (DIC), proposed in Spiegelhalter et al. (2002). Some extensions of DIC to missing data and mixture models have been studied in Celeux et al. (2006). The DIC is defined as a combination of a classical estimate of fit, based on the likelihood function $p(\boldsymbol{Y}|\boldsymbol{\theta})$ and a penalty term, p_D, that represents the complexity of the model. More precisely p_D is defined as

$$p_D = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}} \left[-2 \log p(\boldsymbol{Y}|\boldsymbol{\theta})\right] + 2 \log \left[p(\boldsymbol{Y}|\hat{\boldsymbol{\theta}})\right]$$

in which $\hat{\boldsymbol{\theta}}$ can be the estimated posterior mean, the median or the mode. In our application, the DIC criterion is defined as

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

in which $\hat{\boldsymbol{\theta}}$ is the posterior mean obtained from the MCMC recursions, whereas $\hat{\mathbf{S}}$ is the posterior mode of $p(\mathbf{S}|\boldsymbol{Y})$. It is worth noting that the best model is the one with smaller DIC. We estimated $\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}} \left[-2 \log p(\boldsymbol{Y}|\boldsymbol{\theta}, \mathbf{S}, \boldsymbol{X})\right]$ using the MCMC output, that is,

$$\hat{\mathrm{E}}_{\boldsymbol{\theta}|\boldsymbol{Y}} \left[-2 \log p(\boldsymbol{Y}|\boldsymbol{\theta}, \mathbf{S}, \boldsymbol{X})\right] \approx \frac{-2}{n} \sum_{i=1}^{n} p(\boldsymbol{Y}|\boldsymbol{\theta}^{(i)}, \mathbf{S}^{(i)}, \boldsymbol{X}).$$

6 Simulation Results

In this section we provide some illustrative examples to show the performance of the algorithm. Detailed results of our procedure are reported in Appendix A.

We first simulate a time series of length $T = 1000$ from the ARFIMA$(0, d, 0)$ model

$$(1 - L)^d (y_t - \mu) = \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

and in which the true parameters are $\mu = 1$, $d = 0.4$ and $\sigma^2 = 0.75$. Results are reported in Table 8, based on 10,000 iterations of the MCMC algorithm with a burn-in of 2,500. Since estimates are based on the state space representation provided by equations (7)-(8), our first goal is to determine the truncation parameter M. Chan and Palma (1998) and Ray and Tsay (2002) suggest to use an MA(10) approximation for a time series of length 1000. Our evidence suggests that a reasonable tradeoff between accuracy and computational burden can be based on MA(20) or AR(20) approximations. In particular we refer to a Moving Average approximation since, as suggested in Palma (2007), it should simplify the Kalman recursions. In particular empirical results guarantee accurate estimates of the parameters involved, in particular for the long range persistence d.

We also consider a Monte Carlo analysis for the switching regime long memory model,

$$(1-L)^d(y_t - \mu_{S_t}) = \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2_{S_t}),$$

for different levels of persistence d, ranging between 0.2 and 0.4, keeping the other parameters fixed to, respectively, $\mu_0 = 1$, $\mu_1 = 4$, $\sigma_0^2 = 0.25$, $\sigma_1^2 = 2$, $\pi_{01} = 0.04$ and $\pi_{10} = 0.1$. Posterior's mean results and estimated latent processes are shown in Appendix A, Table 9 to 12 and in Figure 3 to 6. Our findings can be summarized as follows:

- In all the experiments, the estimate of d is accurate, and this is true in particular when the persistence is high. As in the non-switching experiment, a reasonable precision level is achieved for a truncation parameter $M = 20$. We also prefer MA approximations, in line with Palma (2007), since for low levels of persistence, i.e. $d = 0.2$ and $d = 0.3$, we obtain more precise estimates for d.

- For all the simulations, the estimates of the probability of being on state *i=0,1*, namely $\hat{P}(S_t = i)$ are satisfactory, as evidenced in Figure 3 to 6, since they fit accurately the true latent process S_t. In particular this finding is independent on the truncation M and on the different type of approximations (AR or MA).

- On the other side, the estimates of the parameters associated to the high level of volatility, i.e. $S_t = 1$ are imprecise and in particular we underestimate the intercept μ_1 whereas we overestimate σ_1^2. This ill behavior is common for different M and AR or MA approximations. Some further results, not reported here, suggest that this lack of identification is likely due to the small number of contiguous observation labeled by 1. In particular if we consider simulations with more contiguous observations labeled by 1, we obtain more precise estimates also for μ_1 and σ_1^2.

7 Empirical Application: S&P500 realized volatilty

The empirical analysis is based on 5 minutes 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) used to evaluate the likelihood function.

We run the algorithm for 7500 iterations discarding the first 2500. In our experience this choice for the burn-in is more than adequate, even though our aim is to completely remove the effect of the initial values. Results for the model considered are reported in Table 2. An estimate of the switching regime process for the different models is given in Figure 2.

From Table 2 we notice that all the models including VIX outperform the others, in terms of smaller DIC and higher average likelihood, thus stressing the importance of the implied volatility as a predictor. A careful look at Table 2 also evidences that parameters of **Model 1** characterizing conditional means and variances, i.e. μ_i and σ_i^2, are sensibly different across regimes, whereas β_i and d_i are quite stable. These results are consistent with He and Maheu (2009) who find a partial structural break on the NASDAQ volatilities specification in which only the intercept of the conditional variance equation has breaks. For this reason **Model 1** and **Model 1 A** 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 1 B**, provides the better fit in

Table 2: Parameter Estimates of the posterior means. The 95% *credible* intervals are presented between brackets.

	Model 1	Model 1 A	Model 1 B	Model 2	Model 2 A	Model 2 B	Model 3
μ_0	-0.20650	-0.18024	-0.42275	1.27252	1.26608	1.47553	1.18973
	[-0.344,-0.087]	[-0.333,0.005]	[-0.591,-0.241]	[1.153,1.409]	[1.146,1.387]	[1.321,1.587]	[1.158,1.216]
μ_1	0.10496	0.20440		1.74490	1.73002		2.23835
	[-0.146,0.385]	[-0.025,0.504]		[1.542,1.942]	[1.542,1.913]		[2.158,2.295]
β_0	0.90971	0.89549	1.09018				
	[0.832,0.997]	[0.790,0.989]	[0.999,1.179]				
β_1	0.93808						
	[0.816,1.043]						
d_0	0.21511	0.21475	0.25030	0.40409	0.40151	0.44192	
	[0.156,0.277]	[0.170,0.264]	[0.209,0.291]	[0.355,0.459]	[0.360,0.440]	[.400,0.476]	
d_1	0.20724			0.38248			
	[0.093,0.318]			[0.293,0.476]			
σ_0^2	0.08444	0.08471	0.19700	0.09533	0.09580	0.24007	0.12839
	[0.071,0.101]	[0.072,0.101]	[0.182,0.212]	[0.081,0.112]	[0.081,0.114]	[0.217,0.259]	[0.115,0.143]
σ_1^2	0.51279	0.51798		0.58700	0.58626		0.60583
	[0.422,0.628]	[0.427,0.635]		[0.485,0.712]	[0.481,0.718]		[0.531,0.691]
π_{01}	0.02859	0.02918		0.02861	0.02899		0.01619
	[0.016,0.043]	[0.016,0.044]		[0.015,0.045]	[0.016,0.045]		[0.008,0.026]
π_{10}	0.07510	0.07697		0.06585	0.06704		0.02766
	[0.041,0.118]	[0.041,0.121]		[0.035,0.105]	[0.037,0.106]		[0.014,0.044]
DIC	1532.2	1533.4	1515.9	1776.2	1794.1	1782.1	1894.5
log. lik	-699.4	-702.08	-756.00	-811.7	-813.9	-889.5	-928.6

terms of DIC, even though the average likelihood of its switching counterparts are larger. This result can be likely explained by an high penalty factor p_D that account for the superior complexity of **Model 1** and **Model 1 A**. However, on closer inspection, when parameters are free to move across regimes, we obtain sensibly different estimates, thus suggesting some misspecification in **Model 1 B**.

Regarding long memory, 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) and can be easily explained by the impact of VIX that is characterized by strong persistence. It is also confirmed by results on models of class 2, by observing that the persistence parameters almost double, moving from about 0.20 to 0.40. The higher level of persistence d is observed for **Model 2 B**, in which the regime shifts are not considered. This result suggests that the long range persistence absorbs the effect of the nonlinearities characterizing the data. We observe that models of class 1 and 2 behave in a similar way, and in particular it appear that the non-switching models provide a superior fit. However, a more accurate look at Table 2 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 A** and **Model 2 A**. Analysis on **Model 3** evidences that the exclusion of long memory can be also troublesome, since the model provide a poor fit.

Figure 2 evidence that all the models considered provide similar estimates of the regimes, by detecting a systematic 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 now provide and analyze the forecasting performance of the models considered. Our full sample consists on $T = 1241$ observations and we consider a fixed estimation window of $N = 950$. In our analysis, we consider the follow-

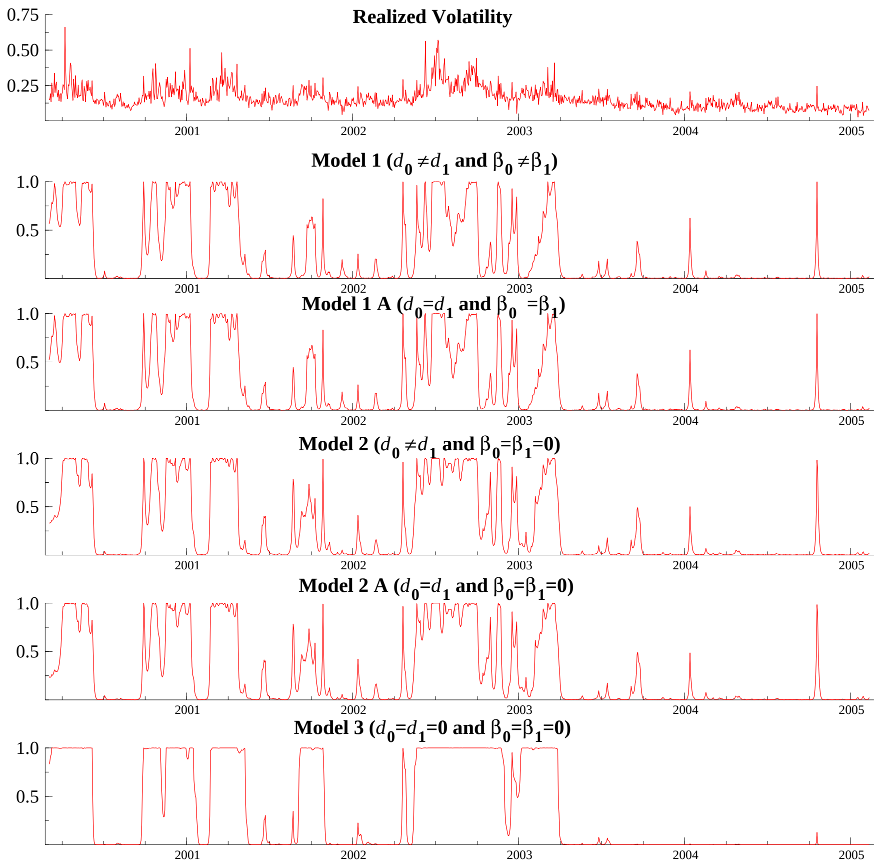

Figure 2: **S&P 500:** Realized volatility and estimated posterior probability of $Pr(S_t = 1|\boldsymbol{Y})$. On the upper panel the *observed* realized volatility. From second to sixth row we report $\widehat{Pr}(S_t = 1|\boldsymbol{Y})$ for all the switching models introduced in Section 3.

ing forecasting horizons, $h = 1, 5, 10$ and 20, corresponding to one trading day, one and two weeks and a trading month, leading to a sequence of, respectively, $K_h = 291, 287, 282, 272$ prediction densities. For the model with the exogenous regressor we compute the forecasts by fixing the future realization of VIX to its actual realization, namely **Mod 1 (fix VIX)**, and forecasting this index by hypothesizing an ARFIMA(0, d_x, 0) model, i.e. **Mod 1 (fore VIX)**. Results on forecasting are reported in Tables from 3 to 7.

In particular Table 3 presents some classic forecasting accuracy indicators, namely RMSE and MAE. There is a clear evidence that models of class 1 provide the better forecasting performance even at longer horizons. Of course, the choice of the future VIX realization as a predictor plays a crucial role in finding these results. In this case VIX seems to be more important than the Markov switching dynamics, as evidenced by the results obtained for **Model 1 B**, that is characterized just by long memory and the implied volatility index. On the other hand, when VIX is not assumed to be known and is evaluated through the scheme presented in Section 4.3, we obtain a general forecasting impoverishment, mainly in the short run. In fact, uncertainties about modeling future VIX sensibly contribute to increase forecasting errors on RV. However, this lack of performance is attenuated in the long run, when the long memory dynamics appear to be more relevant also in the forecast of VIX itself.

Models of class 2 provide satisfactory forecasting performances. In particular results based on **Model 2 A**, that is characterized by long memory and switching regimes, evidence the need to joint modeling nonlinearities and high persistence. The results on **Model 2 A** are valid both in the short as well as in the long run even if, with $h = 20$, **Model 2 B** appear to be superior. This finding suggests that long memory is determinant for longer horizon forecasts. This evidence is also confirmed by the results on the pure switching model, i.e. **Model 3**, for which the RMSE sensibly increase with h.

The results based on RMSE and MAE for different h are also supported by pairwise forecasting comparisons computed through the *Diebold and Mar-*

iano tests[6] (Diebold and Mariano, 1995). We also consider the WLR test of Amisano and Giacomini (2007), as evidenced in Table from 4 to 7 where we report the pairwise comparisons and the associated p-values on parenthesis. A deeper analysis on WLR evidence that VIX is a good predictor of the S&P 500 realized volatility, provided that also VIX is accurately forecasted. Furthermore it is evident that long memory and nonlinearities, even when VIX is not considered, improve the forecasting performance of the proposed model. In particular, according to WLR test, the best model of class 2 is the long memory switching regime with state dependent level, variance and persistency. Furthermore, long memory become important in the long range.

[6]These results are not reported here for conciseness but are available upon request.

Table 3: RMSE and MAE: **1 to 20 Days-Ahead forecasts**.

	1 day-ahead		5 days-ahead		10 days-ahead		20 days-ahead	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Model 1 (fix VIX)	0.24928	0.18433	0.24855	0.18935	0.24979	0.19016	0.25704	0.19725
Model 1 (fore VIX)	0.54577	0.47666	0.37197	0.27845	0.30209	0.22747	0.35506	0.28988
Model 1A (fix VIX)	0.38551	0.30843	0.37293	0.29399	0.38114	0.30662	0.39183	0.30819
Model 1A (fore VIX)	0.47933	0.39580	0.38976	0.30466	0.39384	0.30810	0.48125	0.39666
Model 1B (fix VIX)	0.24916	0.18072	0.24671	0.17861	0.24959	0.18109	0.25480	0.18505
Model 1B (fore VIX)	0.69710	0.64285	0.47065	0.38264	0.34470	0.25784	0.31822	0.25331
Model 2	0.35148	0.29458	0.48026	0.42646	0.59525	0.54504	0.62843	0.57974
Model 2A	0.31725	0.23584	0.29696	0.23116	0.33815	0.27320	0.43919	0.37502
Model 2B	0.48941	0.40926	0.35436	0.26557	0.30528	0.23285	0.37109	0.30199
Model 3	0.43913	0.38303	0.49219	0.43794	0.54756	0.49564	0.62771	0.57786

Table 4: Density Forecast Pairwise comparisons: **1-Day-Ahead forecasts**: (Amisano and Giacomini, 2007) tests and p-values (in parenthesis).

	Alternative Model							
	Mod 1 (fore)	Mod 1A (fix)	Mod 1A (fore)	Model 1B (fix)	Model 1B (fore)	Model 2	Model 2A	Model 2B
Mod 1 (fix VIX)	158.60	109.02	131.98	24.677	53.318	257.59	56.030	44.371
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1 (fore VIX)		-79.185	-190.46	-1784.8	9.2833	-126.94	-68.284	-46.102
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1A (fix VIX)			64.389	-13.034	35.353	-73.009	-17.386	20.261
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1A (fore VIX)				-13619.	16.427	-104.45	-56.476	-36.002
					(0.000)	(0.000)	(0.000)	(0.000)
Mod 1B (fix VIX)					66.595	-4.5812	7.8880	62.167
					(0.000)	(0.000)	(0.000)	(0.000)
Mod 1B (fore VIX)						-45.949	-34.024	-71.261
						(0.000)	(0.000)	(0.000)
Mod 2							30.271	33.769
							(0.000)	(0.000)
Mod 2A								19.968
								(0.000)

Table 5: Density Forecasts Pairwise comparison: **5-Day-Ahead forecasts**: (Amisano and Giacomini, 2007) tests and p-values (in parenthesis).

	Alternative Model							
	Mod 1 (fore)	Mod 1A (fix)	Mod 1A (fore)	Model 1B (fix)	Model 1B (fore)	Model 2	Model 2A	Model 2B
Mod 1 (fix VIX)	63.532	82.683	103.92	11.570	29.375	11.636	23.872	18.740
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1 (fore VIX)		-16.050	115.06	-36.208	12.849	-4.1559	-180.11	2.0459
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.041)
Mod 1A (fix VIX)			34.704	-6.0774	14.185	2.1210	-6.7151	7.0263
			(0.000)	(0.000)	(0.000)	(0.034)	(0.000)	(0.000)
Mod 1A (fore VIX)				-47.642	2.3689	-10.050	-148.44	-5.4867
					(0.018)	(0.000)	(0.000)	(0.000)
Mod 1B (fix VIX)					67.331	11.698	4.8811	28.119
						(0.000)	(0.000)	(0.000)
Mod 1B (fore VIX)						-32.977	-32.886	-34.179
						(0.000)	(0.000)	(0.000)
Mod 2							-6.9091	31.749
							(0.000)	(0.000)
Mod 2A								16.277
								(0.000)

Table 6: Density Forecasts Pairwise comparison: **10-Day-Ahead forecasts**: (Amisano and Giacomini, 2007) tests and p-values (in parenthesis).

	Alternative Model							
	Mod 1 (fore)	Mod 1A (fix)	Mod 1A (fore)	Model 1B (fix)	Model 1B (fore)	Model 2	Model 2A	Model 2B
Mod 1 (fix VIX)	27.675	18.701	33.246	73.840	393.01	13.783	53.071	87.610
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1 (fore VIX)		1.6806	41.681	-12.169	22.616	6.6495	-3.3400	15.282
		(0.0928)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1A (fix VIX)			199.93	-7.8347	13.507	5.9567	-2.4971	11.066
			(0.000)	(0.000)	(0.000)	(0.000)	(0.012)	(0.000)
Mod 1A (fore VIX)				-26.003	-4.9700	1.7172	-24.963	-1.4072
				(0.000)	(0.000)	(0.086)	(0.000)	(0.159)
Mod 1B (fix VIX)					222.27	9.4318	31.121	39.591
					(0.000)	(0.000)	(0.000)	(0.000)
Mod 1B (fore VIX)						3.8043	-58.603	6.2649
						(0.000)	(0.000)	(0.000)
Mod 2							-7.6119	-3.3872
							(0.000)	(0.001)
Mod 2A								24.061
								(0.000)

Table 7: Density Forecasts Pairwise comparison: **20-Day-Ahead forecasts**: (Amisano and Giacomini, 2007) tests and p-values (in parenthesis).

	Alternative Model							
	Mod 1 (fore)	Mod 1A (fix)	Mod 1A (fore)	Model 1B (fix)	Model 1B (fore)	Model 2	Model 2A	Model 2B
Mod 1 (fix VIX)	4.3190	1.5737	13.367	6.7240	4.0974	4.9478	5.8705	6.1766
	(0.000)	(0.115)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1 (fore VIX)		0.56632	3.1426	-0.19860	3.8022	5.2734	11.121	8.2284
		(0.571)	(0.002)	(0.842)	(0.000)	(0.000)	(0.000)	(0.000)
Mod 1A (fix VIX)			0.51536	-0.85313	-0.069307	0.88288	-0.0043422	0.61235
			(0.606)	(0.393)	(0.944)	(0.377)	(0.996)	(0.540)
Mod 1A (fore VIX)				-42.745	-1.0403	1.2965	-1.0615	0.76400
				(0.000)	(0.298)	(0.199)	(0.288)	(0.445)
Mod 1B (fix VIX)					1.3521	2.9775	1.9056	3.1058
					(0.176)	(0.003)	(0.057)	(0.002)
Mod 1B (fore VIX)						6.2099	0.94577	29.764
						(0.000)	(0.344)	(0.000)
Mod 2							-4.2168	-2.6649
							(0.000)	(0.008)
Mod 2A								6.8267
								(0.000)

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.

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.

Brownlees, C., Gallo, G., 2006. Financial Econometric Analysis at Ultra-High Frequency: Data Handling Concerns. Computational Statistics & Data Analysis 51, 2232–2245.

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 Anlysis of Markov Switching Stochastic Volatility Models. Computational Statistics & Data Analysis 51, 4526–4542.

Celeux, G., Forbes, F., Robert, C., Titterington, 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, 719740.

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.

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.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Diebold, F. X., Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13, 253–263.

Doornik, J., 2001. Ox: An Object-Oriented Matrix Programming Language. Timberlake Consultants Press, London.

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.

Hamilton, J., 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica 57, 357–384.

He, Z., Maheu, J. M., 2009. Real Time Detection of Structural Breaks in GARCH Models. Computational Statistics & Data Analysis (forthcoming).

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.

Kim, C. J., Nelson, C., 1999. State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. Review of Economic Studies 65, 361–393.

Koopman, S., Jungnaker, B., Hol, E., 2005. Forecasting Daily Variability of the S&P100 Stock Index Using Historical, Realised and Implied Volatility Mesurements. 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, 113166.

Longin, F., 1997. The Treshold Effect in Expected Volatility: A Model Based on Asymmetric Information. Review of Financial Studies 10, 837–869.

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 Macroeconoomic 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.

Palma, W., 2007. Long-Memory Time Series: Theory and Methods. Wiley Series in Probability and Statistics. J. Wiley & Sons Ed., Hoboken, NJ.

Perron, B., Qu, Z., 2009. Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices. Journal of Business & Economic Statistics forthcoming.

Philippe, A., 1997. Simulation of Right and Left Truncated Gamma Distributions by Mixtures. Statistics and Computing 7, 173–182.

Ray, B. K., Tsay, R. S., 2002. Bayesian Methods for Change-Point Detection in Long-Range Dependent Processes. Journal of Time Series Analysis 23.

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, Ser. B 64, 583–639.

A Monte Carlo Study

A.1 Non-switching ARFIMA model

We first consider a simple long memory model

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

Table 8: **Simulation Results**: Estimates for non-switching long memory model, AR(M) and MA(M) approximations.

	MA(10)	MA(20)	MA(50)	AR(10)	AR(20)	AR(50)
$\mu = 1.00$	1.10778	1.08841	1.06435	1.05364	1.00888	0.91845
	[0.90,1.36]	[0.82,1.33]	[0.72,1.41]	[0.74,1.35]	[0.65,1.34]	[0.46,1.32]
$d = 0.40$	0.36710	0.38143	0.36941	0.41132	0.38971	0.36542
	[0.31,0.44]	[0.32,0.44]	[0.30,0.43]	[0.35,0.47]	[0.33,0.46]	[0.30,0.43]
$\sigma^2 = 0.75$	0.80088	0.77027	0.76209	0.76150	0.73795	0.73263
	[0.70,0.94]	[0.68,0.87]	[0.67,0.86]	[0.67,0.86]	[0.65,1.34]	[0.64,0.82]

A.2 Switching ARFIMA model

Case 1: $d = 0.4$

Table 9: **Simulation Results**: Estimates for switching long memory model, AR(M) and MA(M) approximations.

	MA(10)	MA(20)	MA(50)	AR(10)	AR(20)	AR(50)
$\mu_0 = 1.00$	1.15943	1.19887	1.20683	1.24693	1.25602	1.24462
	[1.060,1.256]	[1.100,1.297]	[1.109,1.304]	[1.139,1.356]	[1.152,1.358]	[1.139,1.349]
$\mu_1 = 4.00$	2.93593	2.77586	2.71657	2.71689	2.67341	2.65610
	[2.635,3.259]	[2.499,3.069]	[2.455,2.994]	[2.444,3.004]	[2.406,2.949]	[2.395,2.920]
$d = 0.40$	0.36935	0.39807	0.41950	0.45200	0.44717	0.45415
	[0.319,0.418]	[0.348,0.450]	[0.360,0.475]	[0.387,0.496]	[0.387,0.495]	[0.393,0.496]
$\sigma_0^2 = 0.25$	0.44633	0.43034	0.41869	0.41835	0.41661	0.42144
	[0.369,0.533]	[0.351,0.522]	[0.339,0.513]	[0.338,0.513]	[0.336,0.508]	[0.340,0.516]
$\sigma_1^2 = 2.00$	2.61606	2.66735	2.63564	2.59117	2.56555	2.56267
	[2.205,3.078]	[2.253,3.143]	[2.241,3.099]	[2.186,3.077]	[2.161,3.041]	[2.173,3.023]
$\pi_{01} = 0.04$	0.04435	0.04745	0.04721	0.04875	0.04858	0.04986
	[0.029,0.062]	[0.031,0.065]	[0.031,0.066]	[0.032,0.067]	[0.032,0.067]	[0.033,0.069]
$\pi_{10} = 0.10$	0.08453	0.0851	0.08183	0.08737	0.08552	0.08517
	[0.055,0.118]	[0.056,0.119]	[0.054,0.114]	[0.058,0.121]	[0.057,0.120]	[0.056,0.120]
DIC	3194.0	3275.7	3241.8	3142.9	3025.7	
log lik.	-1522.6	-1539.9	-1536.9	-1485.48	-1431.8	-1270.4

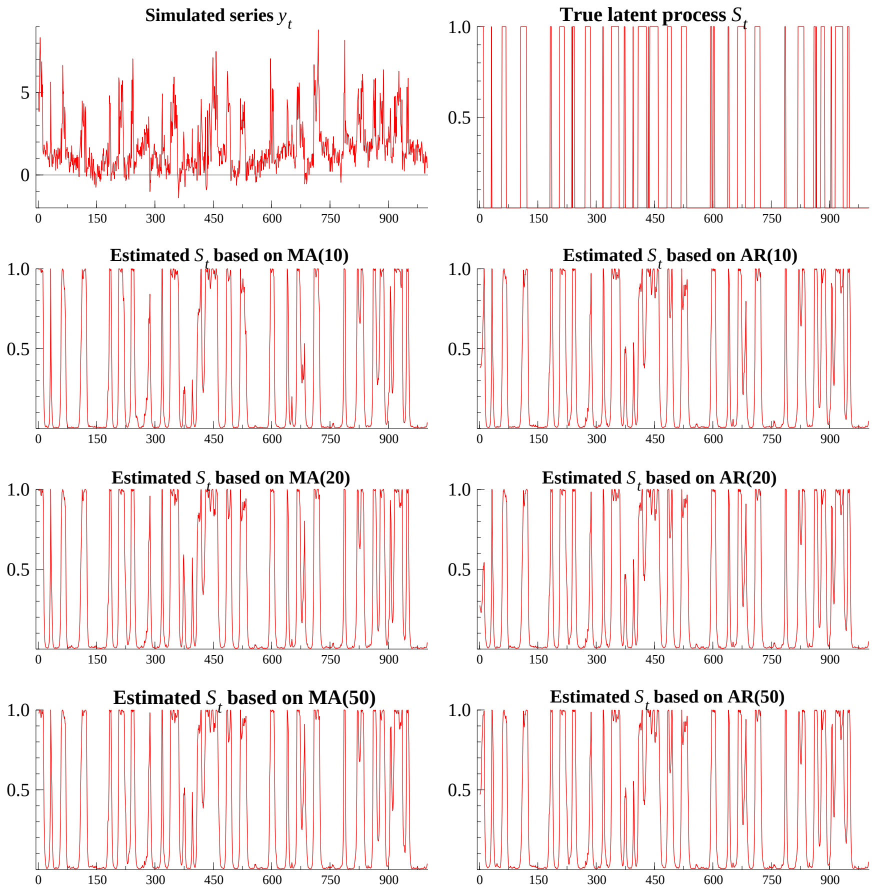

Figure 3: **Simulation Study:** True vs. estimated posterior probability $Pr(S_t = 1|\boldsymbol{Y})$. On the two upper panels the simulated time series y_t and the true (simulated) latent process. On the lower left panels the estimates of $Pr(S_t = 1|\boldsymbol{Y})$ based respectively on the **MA(10)**, **MA(20)** and **MA(50) approximation** whereas on the right side the estimates based on the **AR(10)**, **AR(20)** and **AR(50) approximation**.

Case 1a: $d = 0.4$

Table 10: **Simulation Results**: Estimates for switching long memory model, AR(M) and MA(M) approximations.

	MA(10)	MA(20)	MA(50)	AR(10)	AR(20)	AR(50)
$\mu_0 = 1.00$	0.97493	0.99159	1.01338	1.01569	1.02992	1.03415
	[0.880,1.065]	[0.902,1.083]	[0.927,1.098]	[0.926,1.103]	[0.937,1.123]	[0.941,1.131]
$\mu_1 = 4.00$	2.72005	2.59200	2.54199	2.58324	2.49881	2.48401
	[2.404,3.045]	[2.283,2.913]	[2.252,2.843]	[2.301,2.876]	[2.222,2.798]	[2.207,2.775]
$d = 0.40$	0.34790	0.37333	0.40632	0.40999	0.41413	0.41924
	[0.294,0.399]	[0.311,0.431]	[0.341,0.468]	[0.340,0.480]	[0.346,0.482]	[0.341,0.488]
$\sigma_0^2 = 0.25$	0.39642	0.39143	0.38893	0.38744	0.38714	0.39534
	[0.336,0.467]	[0.325,0.466]	[0.326,0.462]	[0.325,0.460]	[0.324,0.460]	[0.328,0.473]
$\sigma_1^2 = 2.00$	2.99332	2.96153	2.93313	2.87833	2.82423	2.81129
	[2.555,3.505]	[2.544,3.464]	[2.516,3.415]	[2.464,3.357]	[2.416,3.316]	[2.409,3.285]
$\pi_{01} = 0.04$	0.03953	0.04004	0.03946	0.04031	0.04009	0.04050
	[0.025,0.056]	[0.025,0.057]	[0.025,0.057]	[0.025,0.057]	[0.025,0.057]	[0.026,0.058]
$\pi_{10} = 0.10$	0.06714	0.06285	0.06085	0.06450	0.06116	0.06119
	[0.042,0.097]	[0.038,0.092]	[0.037,0.088]	[0.040,0.093]	[0.037,0.089]	[0.037,0.090]
DIC	3103.4	3157.9	3156.4	3006.1	2958.1	2630.2
log lik.	-1498.90	-1511.9	-1512.5	-1459.6	-1411.4	-1247.93

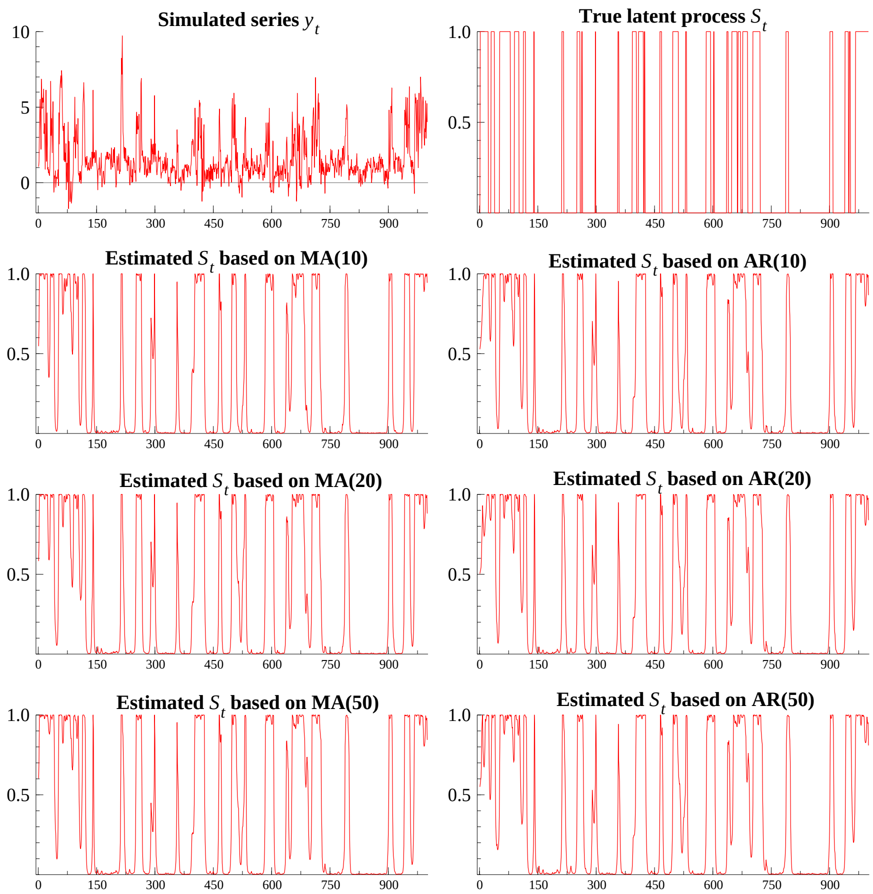

Figure 4: **Simulation Study:** True vs. estimated posterior probability $Pr(S_t = 1|\boldsymbol{Y})$. On the two upper panels the simulated time series y_t and the true (simulated) latent process. On the lower left panels the estimates of $Pr(S_t = 1|\boldsymbol{Y})$ based respectively on the **MA(10)**, **MA(20)** and **MA(50) approximation** whereas on the right side the estimates based on the **AR(10)**, **AR(20)** and **AR(50) approximation**.

Case 2: $d = 0.3$

Table 11: **Simulation Results**: Estimates for switching long memory model, AR(M) and MA(M) approximations.

	MA(10)	MA(20)	MA(50)	AR(10)	AR(20)	AR(50)
$\mu_0 = 1.00$	1.14714	1.14508	1.13702	1.15911	1.14901	1.17972
	[1.047,1.242]	[1.046,1.242]	[1.036,1.239]	[1.062,1.262]	[1.045,1.252]	[1.059,1.300]
$\mu_1 = 4.00$	3.48637	3.31538	3.1059	3.24453	3.12499	3.01907
	[3.135,3.848]	[2.987,3.657]	[2.816,3.413]	[2.944,3.560]	[2.817,3.434]	[2.734,3.329]
$d = 0.30$	0.31519	0.32095	0.36684	0.38767	0.38885	0.37157
	[0.256,0.376]	[0.259,0.382]	[0.298,0.436]	[0.314,0.465]	[0.307,0.465]	[0.293,0.456]
$\sigma_0^2 = 0.25$	0.44889	0.43766	0.40878	0.43316	0.41203	0.40775
	[0.371,0.537]	[0.360,0.524]	[0.330,0.499]	[0.352,0.525]	[0.329,0.499]	[0.324,0.501]
$\sigma_1^2 = 2.00$	2.55694	2.60950	2.57918	2.58254	2.54352	2.53587
	[2.127,3.048]	[2.197,3.081]	[2.210,3.020]	[2.178,3.047]	[2.156,2.975]	[2.164,2.990]
$\pi_{01} = 0.04$	0.03962	0.03951	0.03865	0.04059	0.04001	0.04035
	[0.026,0.055]	[0.025,0.056]	[0.024,0.056]	[0.026,0.057]	[0.025,0.057]	[0.025,0.058]
$\pi_{10} = 0.10$	0.07832	0.07059	0.05991	0.07013	0.06397	0.06254
	[0.051,0.110]	[0.045,0.099]	[0.036,0.088]	[0.045,0.100]	[0.040,0.093]	[0.038,0.093]
DIC	3186.3	3124.5	3107.5	3077.0	2956.5	2558.3
log lik.	-1488.2	-1500.1	-1503.39	-1470.7	-1412.1	-1243.6

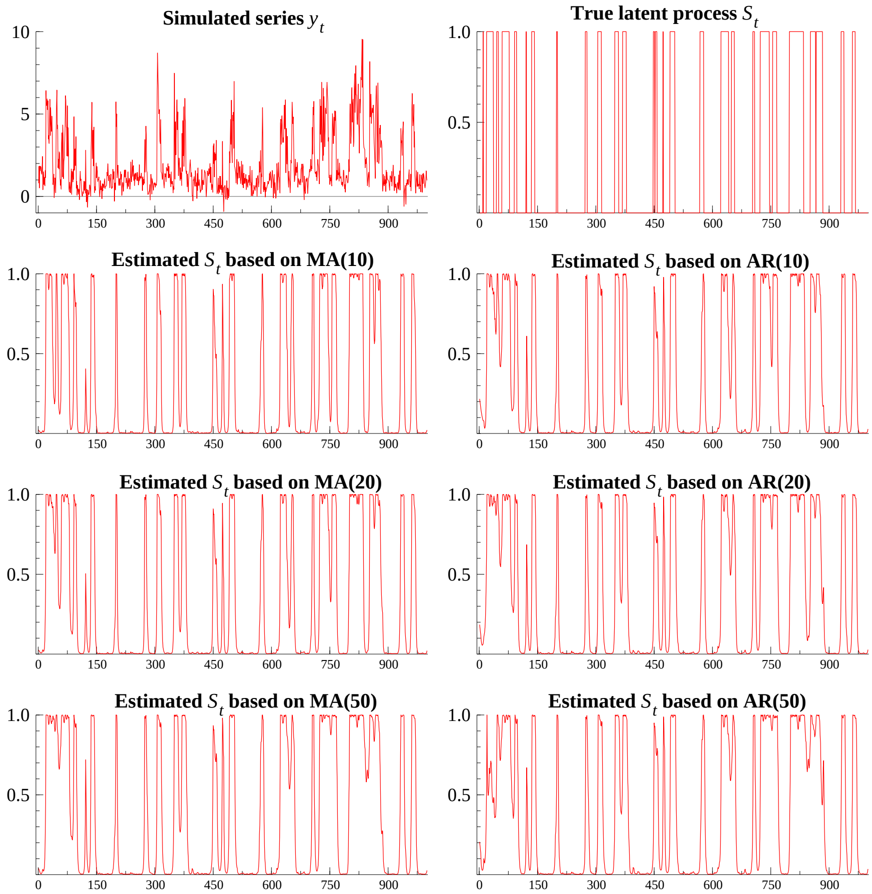

Figure 5: **Simulation Study:** True vs. estimated posterior probability $\widehat{Pr}(S_t = 1|\boldsymbol{Y})$. On the two upper panels the simulated time series y_t and the true (simulated) latent process. On the lower left panels the estimates of $Pr(S_t = 1|\boldsymbol{Y})$ based respectively on the **MA(10)**, **MA(20)** and **MA(50) approximation** whereas on the right side the estimates based on the **AR(10)**, **AR(20)** and **AR(50) approximation**.

Case 3: $d = 0.2$

Table 12: **Simulation Results**: Estimates for switching long memory model, AR(M) and MA(M) approximations.

	MA(10)	MA(20)	MA(50)	AR(10)	AR(20)	AR(50)
$\mu_0 = 1.00$	1.03062	1.12445	1.19104	1.06611	1.19145	1.23316
	[0.699,1.262]	[0.815,1.299]	[0.957,1.339]	[0.714,1.278]	[0.894,1.342]	[1.048,1.361]
$\mu_1 = 4.00$	3.52414	3.34930	3.18270	3.38849	3.23531	3.02713
	[3.084,3.982]	[2.956,3.785]	[2.823,3.587]	[2.967,3.902]	[2.829,3.715]	[2.697,3.412]
$d = 0.20$	0.17883	0.21348	0.24151	0.19933	0.24001	0.28200
	[0.098,0.274]	[0.132,0.294]	[0.158,0.324]	[0.094,0.304]	[0.137,0.336]	[0.180,0.376]
$\sigma_0^2 = 0.25$	0.40635	0.42176	0.44022	0.40820	0.42669	0.44354
	[0.340,0.483]	[0.352,0.501]	[0.364,0.528]	[0.336,0.489]	[0.352,0.510]	[0.365,0.535]
$\sigma_1^2 = 2.00$	2.27277	2.47698	2.59751	2.36020	2.53550	2.59272
	[1.768,2.804]	[2.010,2.975]	[2.182,3.054]	[1.786,2.888]	[2.065,3.009]	[2.196,3.048]
$\pi_{01} = 0.04$	0.04358	0.04402	0.04433	0.04404	0.04269	0.04546
	[0.029,0.059]	[0.029,0.061]	[0.029,0.062]	[0.029,0.060]	[0.028,0.060]	[0.029,0.064]
$\pi_{10} = 0.10$	0.08592	0.08210	0.07773	0.08331	0.07825	0.07412
	[0.058,0.118]	[0.054,0.113]	[0.050,0.108]	[0.056,0.114]	[0.051,0.110]	[0.048,0.104]
DIC	3099.7	3194.7	3236.7	3071.9	3005.0	2770.8
log lik.	-1452.3	-1490.4	-1526.5	-1432.4	-1409.0	-1285.1

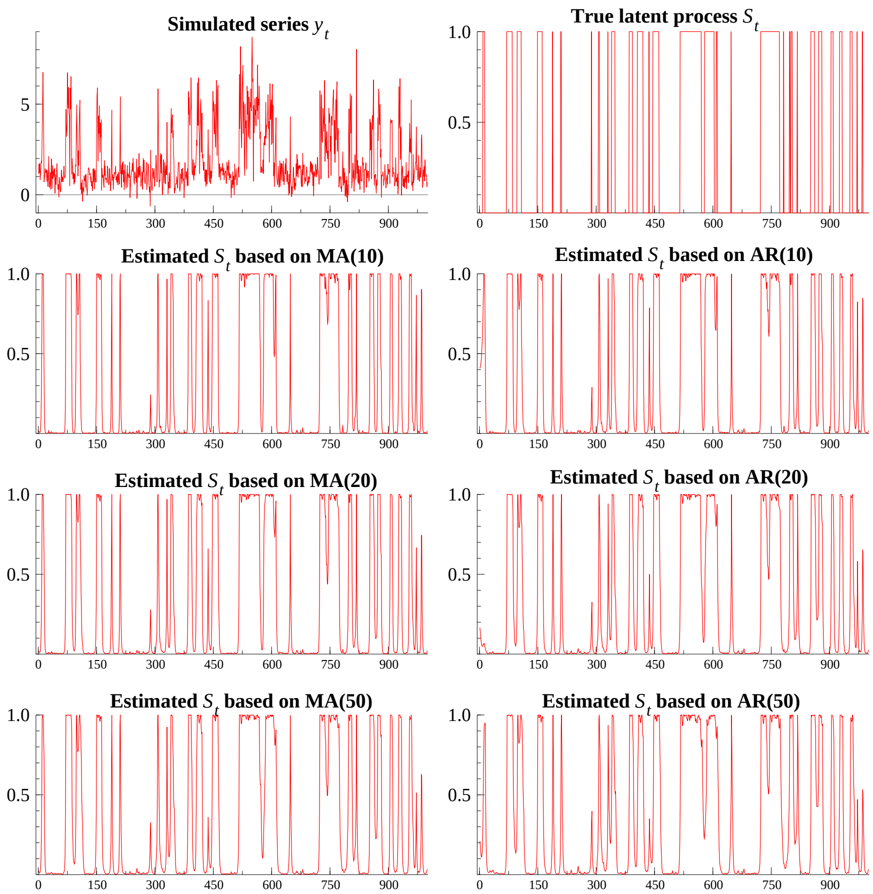

Figure 6: **Simulation Study:** True vs. estimated posterior probability $Pr(S_t = 1|\boldsymbol{Y})$. On the two upper panels the simulated time series y_t and the true (simulated) latent process. On the lower panels the estimates of $Pr(S_t = 1|\boldsymbol{Y})$ based respectively on the **MA(10)**, **MA(20)** and **MA(50) approximation**.