Journal of Computational and Applied Mathematics 300 (2016) 43–55

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Gibbs sampling approach to regime switching analysis of financial time series

Luca Di Persio *, Matteo Frigo

University of Verona, Department of Computer Science, Strada Le Grazie 15, 37100 Verona, Italy

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form 5 December 2015

Keywords:
State-space system
Bayesian analysis
Markov switching
Maximum likelihood
Regime switching
Gibbs sampling

ABSTRACT

We will introduce a Monte Carlo type inference in the framework of Markov Switching models to analyse financial time series, namely the *Gibbs Sampling*. In particular we generalize the results obtained in Albert and Chib (1993), Di Persio and Vettori (2014) and Kim and Nelson (1999) to take into account the switching mean as well as the switching variance case. In particular the volatility of the relevant time series will be treated as a state variable in order to describe the abrupt changes in the behaviour of financial time series which can be implied, e.g., by social, political or economic factors. The accuracy of the proposed analysis will be tested considering financial dataset related to the U.S. stock market in the period 2007–2014.

1. Introduction

Dealing with financial time series is in most cases a non-deterministic task, since a common assumption is the presence of a stochastic error in the empirical datum, and the phenomenon could depend both from observed and unobserved variables, the latter usually called *state variables*. The change of relevant states during time gives rise to the so called *regime switching* dynamic which is governed by specific laws assumed (in literature) to be either deterministic or stochastic.

In this work we are going to deal only with *discrete state-space models* (DSSM) with *Markovian switching* for a specific financial framework taken into account. The aim will be to make a complete quantitative analysis from the rough time series $\{y_t\}_{t=1}^T$, being T a positive integer representing the expiration date or the number of available observations. The resulting system will be of the following form:

$$\begin{cases} y_t = f(S_t, \theta, \psi_{t-1}) \\ S_t = g\left(\tilde{S}_t, \psi_{t-1}\right) \\ S_t \in \Lambda \end{cases} \tag{1}$$

where $\psi_t := \{y_k : k = 1, \dots, t\}$, θ is the vector of the model's parameters, Λ represents the set of the all the possible states, g is the *state-switching* law, namely a function of the past states and the observations until the previous time, while f is the function $f : \Lambda \times \mathbb{R}^k \times \mathbb{R}^{t-1} \to \mathbb{R}$, where k is the number of the *descriptive parameters*, which returns the actual value of the time series at time t. We would like to underline that this class of models is widely used, e.g., in engineering, physics, etc., where they have been implemented to, e.g., study stochastic resonance phenomena, see [1], to underline the relation between corn and oil prices through DSSM, see [2], or to optimal control problem for DC–DC converter systems,

* Corresponding author.
E-mail addresses: luca.dipersio@univr.it (L. Di Persio), matteo.frigo@outlook.com (M. Frigo).

http://dx.doi.org/10.1016/j.cam.2015.12.010
0377-0427/

see [3]. Concerning our analysis, we are going to embed system (1) in the case of a serially uncorrelated time series with Markovian regime switching and a four-dimensional state space. Markov hypothesis is a standard choice in financial time series analysis, while for the choice of the dimensionality we refer to [4], where the authors underline the necessity of distinguishing both the *high risk state* and the *structural break state*. Therefore we obtain the following model:

$$\begin{cases} y_t &= \mu_{S_t} + \epsilon_t, \quad t = 1, 2, \dots, T \\ \epsilon_t &= \text{i.i.d. } \mathcal{N}(0, \sigma^2_{S_t}), \\ \mu_{S_t} &= \mu_1 S_{1,t} + \mu_2 S_{2,t} + \mu_3 S_{3,t} + \mu_4 S_{4,t}, \\ \sigma_{S_t} &= \sigma_1 S_{1,t} + \sigma_2 S_{2,t} + \sigma_3 S_{3,t} + \sigma_4 S_{4,t}, \\ S_t &\in \{1, 2, 3, 4\} \\ p_{ij} &:= \mathbb{P}(S_t = j | S_{t-1} = i) \quad i, j = 1, 2, 3, 4, \end{cases} \tag{2}$$

where $S_{j,t}$ is the characteristic function for the event *being in state j at time t*. We will refer to (2) as to a serially uncorrelated Markov Switching Model (MSM).

Our quantitative study is now shifted to an identification problem with the task of finding the switching probabilities p_{ij} and the *mean–variance* couples $[\mu_m, \sigma_m]$ that describe each state in the DSSM shown in (2). A classical approach to this identification problem is the well known *Maximum Likelihood approach*, which has been exhaustively investigated in [5], to which we refer also for details about the choice of using the *Hamilton filter*.

2. Bayesian inference

Our main problem is to infer on parameters that are subject to stochastic behaviour. In this framework a smart solution is to exploit the *Bayesian Inference* approach, that is a class of methods based on the Bayes' rule. The core of all these methods is the relation between likelihood functions and random variables, namely in order to explain the posterior distribution of a parameter, we write:

$$f\left(\theta|\hat{y}\right) = \frac{f\left(\hat{y}|\theta\right) f\left(\theta\right)}{f\left(\hat{y}\right)}, \tag{3}$$

where on the left hand side we have the joint posterior distribution of the parameters, while on the right hand side we have the product between the likelihood of the data and the prior distribution of the parameters divided by the marginal likelihood of the data (which can be considered constant). The latter suggests to focus the attention on the proportion

$$f\left(\theta|\hat{y}\right) \propto f\left(\hat{y}|\theta\right) f\left(\theta\right). \tag{4}$$

A good choice for the prior distribution would let us compute posterior distributions in an easy way, rather analytically. The latter is not a simple task, but we can exploit the results in [6] to recover that every member of the exponential family has conjugate priors. In particular, if prior and posterior distributions belong to the same family, we say that they are *conjugate distributions*, and the prior is called *conjugate prior for the likelihood*.

2.1. Conjugate distribution

Exploiting the fact that the model is subjected to constraints for the parameters, e.g. we want to have a value for the variance that is not negative, or we would like to preserve the possibility for the mean to take both positive and negative values, we first try to obtain the posterior distribution taken such constraints into account and then we go back to the prior we need to consider. Let us note that every member of the exponential family is endowed with a conjugate prior, hence it is reasonable to exploit such a set, in particular if n is the number of observations with $\tilde{y} = \{y_i\}_{i=1}^n$ being the dataset considered, we have the conjugation properties stated in the following subsections.

2.1.1. Bernoulli with unknown probability p

One basic feature we would like to deal with is the constraint on the inferred parameter that forces it to be a probability value, i.e. $0 \le p \le 1$. A probability distribution that ensures this feature is the *Beta distribution*, that is *self-conjugated*, which means that the prior and the posterior distributions are of the same kind. We can start by considering a random variable $X \sim \text{Bin}(1, p)$, hence X takes values on the set $\{0, 1\}$, $\mathbb{P}(X = 0) = p$ and $\mathbb{P}(X = 1) = 1 - p$. Inferring on p by the Bayesian method means considering p as a random variable and choosing a prior distribution that fits our constraints. One possible easy choice, which is the one we will later adopt, is to suppose that

$$f(p) \propto \text{Beta}(\alpha, \beta), \tag{5}$$

that leads to the posterior distribution of p given $\tilde{y}$:

$$f(p|\tilde{y}) = \text{Beta}\left(\alpha + \sum_{i=1}^n y_i, \beta + n - \sum_{i=1}^n y_i\right). \tag{6}$$

It is worth to mention that the relevance of the previously sketched framework will be clear in Section 3.1.1.

2.1.2. Normal with known variance

Another classical result in Bayesian decision theory is the posterior distribution of a normal random variable with known variance. In this case $X \sim \mathrm{N}(\mu, \sigma^2)$ and we consider σ^2 as a known constant. Since our aim is to make inference on μ, we want to find a prior distribution with the following features:

- Values must be finite, i.e. $|\mu| < \infty \Leftrightarrow \lim_{t\to\infty} f(\mu > t) = 0$,
- It must be symmetric with respect to its mean value.

Latter constraints lead to choose a Normal prior distribution even if this is not the only possible choice. In fact the same properties are satisfied, e.g., by the *Cauchy's distribution*, by the *Student's t-distribution* or by the *logistic distribution*, nevertheless the Normal distribution choice is rather standard because of the Gaussian nature of the likelihood functional, namely

$$f(\mu) = \mathrm{N}(\mu_0, \sigma_0^2).$$

The posterior distribution of μ given $\tilde{y}$ and σ^2 is

$$f(\mu|\tilde{y}, \sigma^2) = \mathrm{N}\left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} y_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)\right). \tag{7}$$

The further equation plays the key role in the generalization of the pure switching variance model proposed by Di Persio and Vettori in [4] and we please you to wait until Section 3.1.3 in order to go into the details of the switching mean MCMC simulation.

2.1.3. Normal with known mean

The complement of the previous result is the inference on σ^2 with known μ in the normal distribution framework. The new property we want to satisfy is $0 < \sigma^2 < +\infty$. We reject the guess made in Section 2.1.2, since the normal distribution takes both positive and negative values and, accordingly to such a restriction, we suppose

$$f(\sigma^2) = \mathrm{IG}(\alpha, \beta),$$

where IG indicates the *inverse-Gamma distribution*, i.e. $\frac{1}{\sigma^2} \sim \Gamma\left(\alpha, \frac{1}{\beta}\right)$. Then the posterior distribution is:

$$f(\sigma^2|\mu, \tilde{y}) \sim \mathrm{IG}\left(\alpha + \frac{n}{2}, \frac{\sum_{i=1}^{n}(y_i + \mu)^2}{2}\right). \tag{8}$$

Even in this case we can make different choices, since the same property is satisfied by the *scale-inverse-chi-squared* distribution, but it is possible to show that the inverse-Gamma distribution is a reparametrization of scale-inverse-chi-square, see, e.g., [7, Example 2.8].

2.2. Gibbs sampling

The system we are studying involves a phenomenon subject to a Markov Chain dynamic, so we would like to take a sampling algorithm from the *Markov Chain Monte Carlo* (MCMC) family. As proposed in [8], we exploit the Gibbs Sampler in order to obtain a suitable sample for the parameters we want to estimate.

Let z_t be the set of the parameters we want to estimate through our MCMC algorithm. In order to start the Gibbs Sampling algorithm we need to know every conditional density $f(z_m|z_{n\neq m})$, where

$$z_{n\neq m} = \{z_1, \dots, z_{m-1}, z_{m+1}, \dots, z_k\},$$

is the set of all the parameters except the mth. The sampler operates through the following steps:

Step 1 Set an initial value z_m^0 for each parameter to be sampled and a counter $c = 1$.
Step 2 Draw every parameter from its conditional distribution through the following procedure:

1. Draw z_1^c from $f(z_1|z_2^{c-1}, \dots, z_k^{c-1})$.[1]
2. Draw z_2^c from $f(z_2|z_1^{c-1}, z_3^{c-1}, \dots, z_k^{c-1})$.

[1] It should be clear that an initial guess for z_1 is not necessary, but it is highly suggested in order to have a clearer algorithm and code.

3. ...
4. Draw z_k^c from $f(z_K|z_1^{c-1}, \ldots, z_{k-1}^{c-1})$.

Step 3 Update the counter $c \Leftarrow c+1$ and restart from Step 2 until $c < J$, where J is number of iterations we want to compute.

The choice of J is one of the key points in the set-up of the Gibbs Sampler, since it needs to be big enough to ensure a good approximation of both the marginal and joint distributions of the sampled values to the marginal and joint distributions of the data. In [9, Appendix], the authors proved this convergence (with exponential rate) as $J \to \infty$. Usually in Monte Carlo methods we need a *burn-in period* to enter in the convergence region, i.e. we have to reject the first L samples, where L differs from problem to problem, so we will take $J = L + M$ where M is the dimension of the sample that will be taken into account for the Monte Carlo estimation of the parameters of interest. We would like to underline that even if the MSM framework is not endowed with the conditional distributions, the Bayesian Inference will let us exploit the Gibbs Sampler in a proper way.

3. Markov switching models inference with Gibbs sampling

The main problem in dealing with Markov Switching Models is that their behaviour is strongly dependent on the unobserved variable S_t. Exploiting previous considerations on Bayesian inference and according to the algorithms presented in [8, Sec. 3] and in [5, Ch. 9], we can consider S_t as a random variable by means of the following considerations. First, in order to set the problem in a more convenient scheme, we rewrite the standard deviation equation of the MSM as follows:

$$\sigma_{S_t}^2 = \sigma_1^2 S_{1,t} + \sigma_2 S_{2,t} + \sigma_3 S_{3,t} + \sigma_4 S_{4,t} \tag{9}$$

$$= \sigma_1^2(1 + h_2 S_{2,t}) \left[(1 + h_2 S_{3,t})(1 + h_3 S_{3,t})\right] \left[(1 + h_2 S_{4,t})(1 + h_3 S_{4,t})(1 + h_4 S_{4,t})\right], \tag{10}$$

with the aim of inferring on the parameter set

$$\Theta = \left\{\mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, h_2, h_3, h_4, \tilde{S}_T, P\right\},$$

where $\tilde{S}_T = \{S_1, S_2, \ldots, S_T\}$ and

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}.$$

Then, in order to apply the Gibbs Sampling, we consider a slightly more general version of the procedure described in [4, Section 3.3], namely:

Step 1 Derive the distribution of S_t, $t = 1, \ldots, T$ conditional on the parameters in one of the following ways:

1. *Single-Move Gibbs Sampling*: generate each S_t from

$$f\left(S_t|\tilde{S}_{\neq t}, \mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, h_2, h_3, h_4, P, \psi_T\right),$$

for $t = 1, \ldots, T$, where $\tilde{S}_{\neq t} = \tilde{S}_T \setminus S_t$.

2. *Multi-Move Gibbs Sampling*: generate the whole block $\tilde{S}_T$ from

$$f\left(\tilde{S}_T|\mu_1, \mu_2, \mu_3, \mu_4, \sigma_1^2, h_2, h_3, h_4, P, \psi_T\right).$$

Step 2 Generate the transition probabilities $p_{i,j}$ from $f(P|\tilde{S}_T)$. Notice that this distribution depends only on $\tilde{S}_T$ because we assume to deal with a homogeneous Markov chain. If we choose the *Beta distribution* as the prior distribution for P, we have that the posterior distribution $f(P|\tilde{S}_T) = f(P)L(P|\tilde{S}_T)$ is again a Beta distribution, hence the Beta distribution is a conjugate prior of the likelihood of transition probabilities, as seen in Section 2.1.1.

Step 3 Generate $\mu_1, \mu_2, \mu_3, \mu_4$ from

$$f\left(\mu_1, \mu_2, \mu_3, \mu_4|\tilde{S}_T, \sigma_1, h_2, h_3, h_4, P, \psi_T\right),$$

in this case the conjugate prior is the *Normal distribution*, see Section 2.1.2 for details.

Step 4 Generate σ_1, h_2, h_3, h_4 from

$$f\left(\sigma_1, h_2, h_3, h_4|\tilde{S}_T, \mu_1, \mu_2, \mu_3, \mu_4, P, \psi_T\right).$$

By (9) we consider, e.g., $\sigma_3^2 = \sigma_1^2(1 + h_2)(1 + h_3)$: we can first generate σ_1^2 from h_2 and h_3, then we generate both $\bar{h}_2 = 1 + h_2$ from σ_1^2 and h_3, and we conclude obtaining $\bar{h}_3 = 1 + h_3$ from σ_1^2 and h_2. In every case we use the *Inverted Gamma distribution* as the conjugate prior for the parameters.

Further results can be obtained turning from the *Single-Move Gibbs Sampling* approach, to the *Multi-Move Gibbs Sampling*, originally motivated in [10] and in [5]. In particular the following section provides details of the embedded version of the Gibbs Sampling with respect to our problem.

3.1. Embedded Gibbs sampling

Let $J = L + M$ be the quantity of samples we want to compute, with L as the *burn-in* period and M the dimension of the inferring sample space, then let $c = 1$ and *compute* as described in the following subsection, until $c < J$.

3.1.1. Multi-move Gibbs sampling for drawing states

Make inference on the state variables conditional on the parameters and the data.

1. Perform the Hamilton's filtering procedure used for the Maximum Likelihood approach, see [11, Sec. 3], in order to obtain
$$\mathbb{P}\left(S_T|\psi_T\right),$$
2. consider the property
$$p_t^i = \mathbb{P}\left(S_t = i|S_{t+1}, \psi_t\right) = \frac{f\left(S_{t+1}|S_t = i\right) f\left(S_t = i|\psi_t\right)}{\sum_{k=1}^{4} f\left(S_{t+1}|S_t = k\right) f\left(S_t = k|\psi_t\right)},$$
and declare the vector
$$\pi_t = \left[p_t^1, p_t^2, p_t^3, p_t^4\right]. \tag{11}$$
3. Simulate a value for S_t from π_t for all $t = 1, \ldots, T$ and save those values in the vector $\tilde{S}_T$.

3.1.2. Generate transition probabilities

Since we want to obtain a set of values that belong to the interval $[0, 1]$, then we recall the analysis reported in 2.1.1 and we exploit the properties of the *Beta distribution*, which implies that we first need to define the following quantities

$$\bar{P} = \begin{cases} \bar{p}_{ij} = 1 - p_{ii}, & \text{if } i = j, \\ \bar{p}_{ij} = p_{ij}, & \text{if } i \neq j, \end{cases} \tag{12}$$

therefore we have the following *bar-shifting* property:

$$\begin{aligned} p_{ij} =& \mathbb{P}\left(S_t = j|S_{t-1} = i\right) \\ =& \mathbb{P}\left(S_t = j|S_{t-1} = i, S_t \neq i\right) \mathbb{P}\left(S_t \neq i|S_{t-1} = i\right) \\ =& \bar{p}_{ij}(1 - p_{ii}). \end{aligned} \tag{13}$$

Moreover we need to define the *transition counter*, i.e. a 4-by-4 matrix n_{ij} where each entry counts the number of transitions from state $S_{t-1} = i$ to state $S_t = j$. The last object is the complementary of n_{ij}, namely we set $\bar{n}_{ij}$ be the 4-by-4 matrix where each entry counts the number of transition from state $S_{t-1} = i$ to state $S_t \neq j$. Since we choose the *Beta distribution* as the conjugate prior distribution, we have

$$p_{ij} \sim \text{Beta}\left(u_{ij}, \bar{u}_{ij}\right),$$

and it implies

$$\begin{aligned} p_{ii}|\tilde{S}_T \sim& \text{Beta}\left(u_{ii} + n_{ii}, \bar{u}_{ii} + \bar{n}_{ii}\right), \\ \bar{p}_{ij}|\tilde{S}_T \sim& \text{Beta}\left(u_{ij} + n_{ij}, \bar{u}_{ij} + \bar{n}_{ij}\right). \end{aligned}$$

Then we can compute the transition probabilities row by row. The computation for the first row of the matrix is shown and the others follow the same procedure. We start by obtaining the element from the diagonal p_{11}, then generating a random value for the other two elements, e.g. $\bar{p}_{12}$ and $\bar{p}_{13}$. We transform these last two values with the *bar-shifting* property of Eq. (13) in order to obtain

$$\begin{aligned} p_{12} &= \bar{p}_{12}\left(1 - p_{11}\right), \\ p_{13} &= \bar{p}_{13}\left(1 - p_{11}\right), \end{aligned}$$

and the last evaluation for this row is $p_{14} = 1 - p_{11} - p_{12} - p_{13}$.

Repeating previous steps for each row we obtain the full specification of the P matrix.

3.1.3. Generate means

Concerning the mean values, we start by considering the subsequences of y_t belonging to a specific state. Let us define $Y^j := \{y_t : S_t = j\}$ and $L^j = \#Y^j$. As long as we are in the same framework of Section 2.1.2, we choose the *Normal distribution* as the conjugate prior distribution for the parameters $\mu_1, \dots, \mu_4$, i.e. we take

$$\mu \sim \mathrm{N}(\mu_0, \sigma_0^2),$$

this leads to the posterior distribution

$$f(\mu_j|Y^j, \sigma_j^2) = \mathrm{N}\left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{L^j} Y_i^j}{\sigma_j^2}}{\frac{1}{\sigma_0^2} + \frac{L^j}{\sigma_j^2}}, \left(\frac{1}{\sigma_0^2} + \frac{L^j}{\sigma_j^2}\right)\right), \tag{14}$$

and we only need to compute a value from Eq. (14) for each state $j = 1, 2, 3, 4$.

3.1.4. Generate variances

Concerning the computation of variances we have to face a rather complicated problem, since there are four parameters to be drawn from an *Inverse Gamma distribution*. Associated computational steps are reported in what follows:

Infer on σ_1^2 conditional to h_2, h_3, h_4. Let us consider the following time series:

$$Y_t^1 := \frac{y_t - \mu_{S_t}}{\sqrt{(1 + h_2 S_{2,t})(1 + h_2 S_{3,t})(1 + h_3 S_{3,t})(1 + h_2 S_{4,t})(1 + h_3 S_{4,t})(1 + h_4 S_{4,t})}},$$

that is equivalent to considering the original time series y_t and performing the following rescaling:

- if $S_t = 1$, $y_t \to y_t - \mu_1$;
- if $S_t = 2$, $y_t \to \frac{y_t - \mu_2}{\sqrt{1+h_2}}$;
- if $S_t = 3$, $y_t \to \frac{y_t - \mu_3}{\sqrt{(1+h_2)(1+h_3)}}$;
- if $S_t = 4$, $y_t \to \frac{y_t - \mu_4}{\sqrt{(1+h_2)(1+h_3)(1+h_4)}}$;

notice that we have $Y_t^1 \sim \mathrm{N}(0, \sigma_1^2)$, then let $L_1 = \#Y_t^1$, hence choosing an *Inverse-Gamma* conjugate prior distribution for σ_1^2, i.e.

$$f\left(\sigma_1^2\right) \sim \mathrm{IG}\left(\frac{\nu_1}{2}, \frac{\delta_1}{2}\right),$$

we can exploit the properties seen in Section 2.1.3 to compute a value for $\frac{1}{\sigma_1^2}$ from the posterior distribution

$$f\left(\frac{1}{\sigma_1^2}|h_2, h_3, h_4\right) = \Gamma\left(\frac{\nu_1 + L_1}{2}, \left[\frac{\delta_1 + \sum_{t=1}^{L_1} Y_t^1}{2}\right]^{-1}\right). \tag{15}$$

Infer on h_2 conditional to σ_1^2, h_3, h_4. Let us define the time series

$$y_t^{(2)} = \{y_t | S_t \in \{2, 3, 4\}\}, \tag{16}$$

and let $L_2 = \#y_t^{(2)}$, then we define Y_t^2 by the following reparametrization of $y_t^{(2)}$:

- if $S_t = 2$, $y_t \to \frac{y_t - \mu_2}{\sqrt{\sigma_1^2}}$;
- if $S_t = 3$, $y_t \to \frac{y_t - \mu_3}{\sqrt{\sigma_1^2(1+h_3)}}$;
- if $S_t = 4$, $y_t \to \frac{y_t - \mu_4}{\sqrt{\sigma_1^2(1+h_3)(1+h_4)}}$.

Let $\bar{h}_2 := 1 + h_2$, then, exploiting the prior distribution, we have

$$f\left(\bar{h}_2\right) \sim \mathrm{IG}\left(\frac{\nu_2}{2}, \frac{\delta_2}{2}\right),$$

and we obtain the following posterior distribution:

$$f\left(\frac{1}{\bar{h}_2}|\sigma_1^2, h_3, h_4\right) = \Gamma\left(\frac{\nu_2 + L_2}{2}, \left[\frac{\delta_2 + \sum_{t=1}^{L_2} Y_t^2}{2}\right]^{-1}\right). \tag{17}$$

Then we compute a value for $\bar{h}_2$ from (17) and, if the drawn value is less than one, simulate (17) until it returns a value $\bar{h}_2 > 1$.
Infer on h_3 conditional to σ_1^2, h_2, h_4. Let us define the time series

$$y_t^{(3)} = \{y_t | S_t \in \{3, 4\}\}, \tag{18}$$

and $L_3 = \#y_t^{(3)}$, then we define Y_t^3 through the following reparametrization of $y_t^{(3)}$:

- if $S_t = 3$, $y_t \to \frac{y_t - \mu_3}{\sqrt{\sigma_1^2(1+h_2)}}$;
- if $S_t = 4$, $y_t \to \frac{y_t - \mu_4}{\sqrt{\sigma_1^2(1+h_2)(1+h_4)}}$.

Let $\bar{h}_3 := 1 + h_3$, from the prior distribution

$$f\left(\bar{h}_3\right) \sim \mathrm{IG}\left(\frac{\nu_3}{2}, \frac{\delta_3}{2}\right),$$

we obtain the following posterior distribution:

$$f\left(\frac{1}{\bar{h}_3}|\sigma_1^2, h_2, h_4\right) = \Gamma\left(\frac{\nu_3 + L_3}{2}, \left[\frac{\delta_3 + \sum_{t=1}^{L_3} Y_t^3}{2}\right]^{-1}\right). \tag{19}$$

Draw a value for $\bar{h}_3$ from (19) and, if it is less than one, simulate (19) until it returns a value $\bar{h}_3 > 1$.
Infer on h_4 conditional to σ_1^2, h_2, h_3. Let us define the time series

$$y_t^{(4)} = \{y_t | S_t = 4\}, \tag{20}$$

and $L_4 = \#y_t^{(4)}$, then we define Y_t^4 through the following reparametrization of $y_t^{(4)}$:

- if $S_t = 4$, $y_t \to \frac{y_t - \mu_4}{\sqrt{\sigma_1^2(1+h_2)(1+h_3)}}$.

Let $\bar{h}_4 := 1 + h_4$, using the prior distribution

$$f\left(\bar{h}_4\right) \sim \mathrm{IG}\left(\frac{\nu_4}{2}, \frac{\delta_4}{2}\right),$$

we obtain the following posterior distribution:

$$f\left(\frac{1}{\bar{h}_4}|\sigma_1^2, h_2, h_3\right) = \Gamma\left(\frac{\nu_4 + L_4}{2}, \left[\frac{\delta_4 + \sum_{t=1}^{L_4} Y_t^4}{2}\right]^{-1}\right). \tag{21}$$

Then we compute a value for $\bar{h}_4$ from (21) and, if it is less than one, we simulate (21) until it returns a value $\bar{h}_4 > 1$.

3.1.5. Update counter

Once the computations stated in subsections from (Step 1) to (Step 4), are completed, we update the counter $c \Leftarrow c + 1$ and restart from Section 3.1.1. If the maximum number of iterations is reached we simply avoid restarting the cycle and the whole procedure ends.

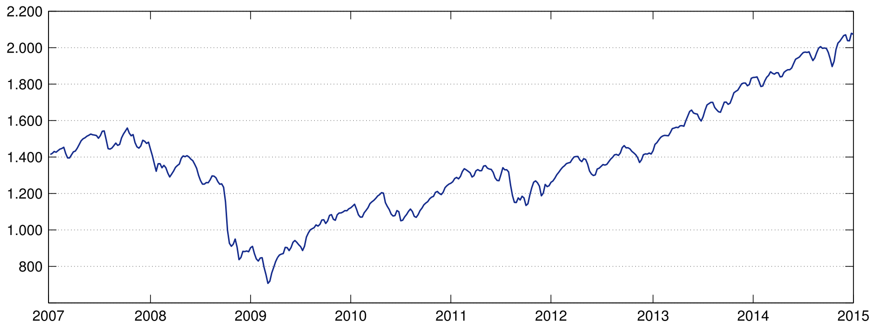


Fig. 1. Stock price of S&P500 defined as the weekly average of the opening and closing prices. The length of the time series is $N = 419$.

4. Goodness of fit

After the identification of the system parameters, obtained either through a Maximum Likelihood or via Gibbs Sampling, we have the transition matrix, the model parameters

$$\theta = \left[\hat{\mu}_1, \ldots, \hat{\mu}_4, \hat{\sigma}_1, \ldots, \hat{\sigma}_4\right],$$

and the filtered probabilities and the smoothed probabilities

$$\mathbb{P}\left(S_t = j|\psi_t\right) \quad \text{and} \quad \mathbb{P}\left(S_t = j|\psi_T\right). \tag{22}$$

In order to estimate the mean and the variance of the process at time t conditional w.r.t. the set ψ_t, we compute the following weighted averages $\forall t = 1, \ldots, T$:

$$\hat{\mu}_t = \mathbb{E}\left[\mu_t|\psi_t\right] = \hat{\mu}_1\mathbb{P}\left(S_t = 1|\psi_t\right) + \cdots + \hat{\mu}_4\mathbb{P}\left(S_t = 4|\psi_t\right) \tag{23}$$

$$\hat{\sigma}_t = \mathbb{E}\left[\sigma_t|\psi_t\right] = \hat{\sigma}_1\mathbb{P}\left(S_t = 1|\psi_t\right) + \cdots + \hat{\sigma}_4\mathbb{P}\left(S_t = 4|\psi_t\right), \tag{24}$$

then we can define the standardized residuals in the following way

$$\hat{\epsilon}_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t} \sim \mathcal{N}\left(0, 1\right), \quad t = 1, \ldots, T, \tag{25}$$

hence obtaining a good benchmark for the regression accuracy. In particular, the *goodness of the fit* is tested by a normality test, which consists of applying a statistical hypothesis test under the bilateral null hypothesis H_0 that the residuals are distributed like $\mathcal{N}(\mu, \sigma^2)$ random samples. We would like to underline that the chosen tests are the ones proposed by Jarque and Bera, see [12], and the one exploited by Lilliefors in [13]. We underline that, as pointed out by Thadewald and Büning in [14], the Jarque–Bera test is sometimes subject to errors of the second type.

Another informative, though qualitative, test is the *Normal Probability Plot*, which compares the percentiles of the standardized residuals with the theoretical values. In this type of plot most of the values should stay close to a particular line, otherwise it is reasonable to conjecture non-normality for the analysed values.

5. Case study: S&P500

In what follows we apply the previously presented procedure to a concrete case study of particular financial relevance. In particular, we present a MSM-inference approach focused on the Standard & Poor's 500 (S&P 500) equity index.

5.1. Long term: 2007–2014

We will start by analysing the period between January 1st 2007 and December 31st 2014 in order to include both the great financial crisis and the economic recovery of the U.S. economy. Fig. 1 shows the weekly stock price of S&P500 index in the chosen time interval defined in the following way:

$$\text{Price}(t) = \frac{\text{Close}(t) - \text{Open}(t)}{2}.$$

Let $\{X_t\}$ be the price time series, then we can define the return process as $y_t = \frac{X_{t+1} - X_t}{X_t}$. From now on we will work only on y_t, whose length is $T = 418$. Fig. 2 shows the weekly return of S&P500 index in the chosen time interval. The first step

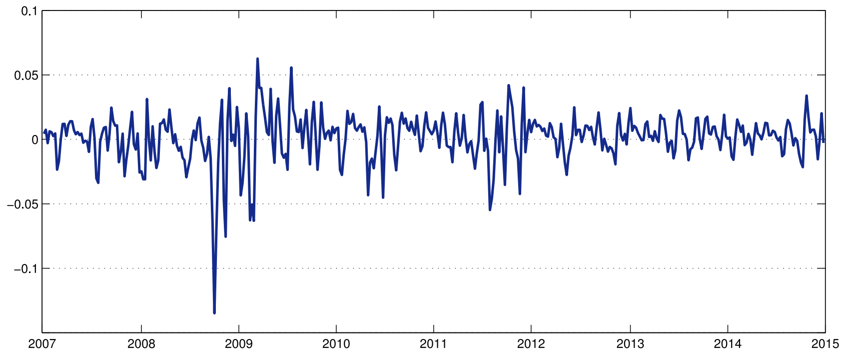


Fig. 2. Return process of S&P500. The length of the time series is $N = 419$.

of our analysis consists in trying to check the possibility for the returns to be autocorrelated, a task which can be accomplished through tests like Durbin–Watson test, see, e.g., [15–17], by directly computing the AR(1) coefficient exploiting the least-square methods in the form

$$\phi_1 = \frac{\sum_{i=2}^{T} y_{t-1} y_t}{\sum_{i=1}^{T-1} (y_t)^2}.$$

In our case this computation returns the value −0.0489, which is low enough to discard the assumption of a first-order autoregressive pattern, hence we decide to focus our analysis on a Gaussian distribution with switching variance considering the following four-states model:

$$y_t = \epsilon_t, \quad t = 1, 2, \dots, T \tag{26}$$

$$\epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{S_t}^2), \tag{27}$$

$$\sigma_{S_t} = \sigma_1 S_{1,t} + \sigma_2 S_{2,t} + \sigma_3 S_{3,t} + \sigma_4 S_{4,t}, \tag{28}$$

$$p_{ij} := \mathbb{P}(S_t = j | S_{t-1} = i) \quad i, j = 1, 2, 3, 4, \tag{29}$$

where $S_{j,t}$ is the characteristic function for the event *being in state j at time t* and the four states correspond respectively to the low volatility, medium volatility, high volatility and very high volatility regimes. We underline that the possibility of a switching mean model has not been taken into account because it is not informative for the analysed time series which, in fact, does not show evidence of sensible non-zero-mean state presence.

Exploiting the Hamilton filter for the return process, we obtain the results shown in Figs. 3 and 4 with the following parameters:

$$P = \begin{bmatrix} 0.9779 & 0.0176 & 0.0000 & 0.0045 \\ 0.0417 & 0.9583 & 0.0000 & 0.0000 \\ 0.0000 & 0.0377 & 0.9622 & 0.0001 \\ 0.0004 & 0.0004 & 0.2990 & 0.7002 \end{bmatrix},$$

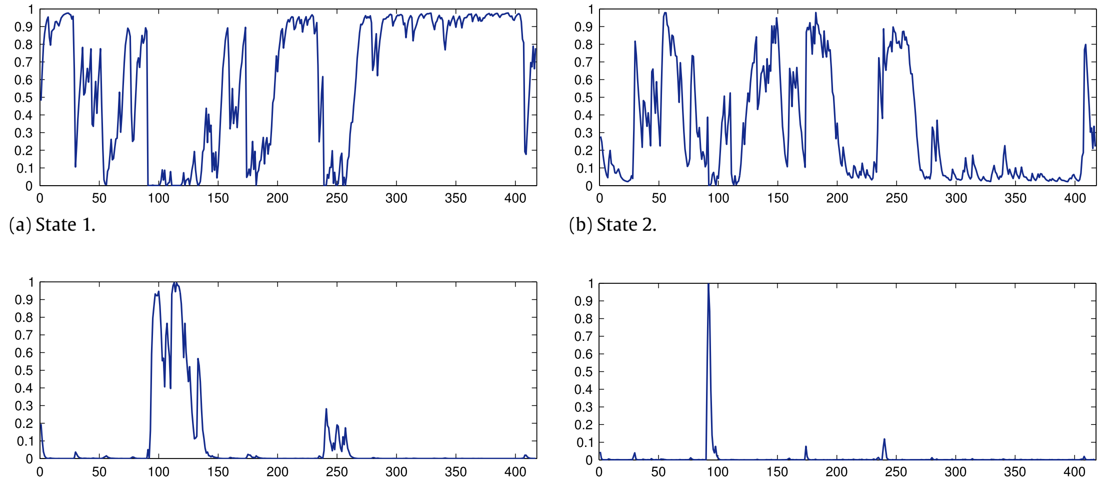
$$\sigma_1^2 = 0.0001, \qquad \sigma_2^2 = 0.0004, \qquad \sigma_3^2 = 0.0013, \qquad \sigma_4^2 = 0.0073.$$

Considering the standard deviation

$$\hat{\sigma}_t = \mathbb{E}(\sigma_t | \psi_t) = \hat{\sigma}_1 \mathbb{P}(S_t = 1 | \psi_t) + \cdots + \hat{\sigma}_4 \mathbb{P}(S_t = 4 | \psi_t), \quad t \in [1, T],$$

which we would like to compare with the VIX, namely the *Chicago Board Options Exchange Market Volatility Index* (CBOE VIX), which is a widely used measure of the implied volatility of the S&P500 index. Let us underline that speaking of implied volatility, we mean the expectation of future volatility. Fig. 5 shows a graphical comparison between our estimate and the VIX index.

Let us state one consideration on this heuristic comparison: first we can say that the initial period, which approximately, corresponds to the first fifty observations, is not a reliable estimate for the standard deviation since it is really far from VIX and has the typical behaviour of a *warm-up* period. The second consideration is that it is possible to notice a smoother pattern for the estimated standard deviation. The latter leads to an underestimate for all the peaks of volatility even if they are reached at the same time. The only peak that is overestimated with respect to the VIX is the highest for both the time series, corresponding to the recession experienced in 2008. Underlining that previous facts turn to be hold true only in chaotic periods, our estimate is a robust and efficient *substitute* for the volatility of S&P500.



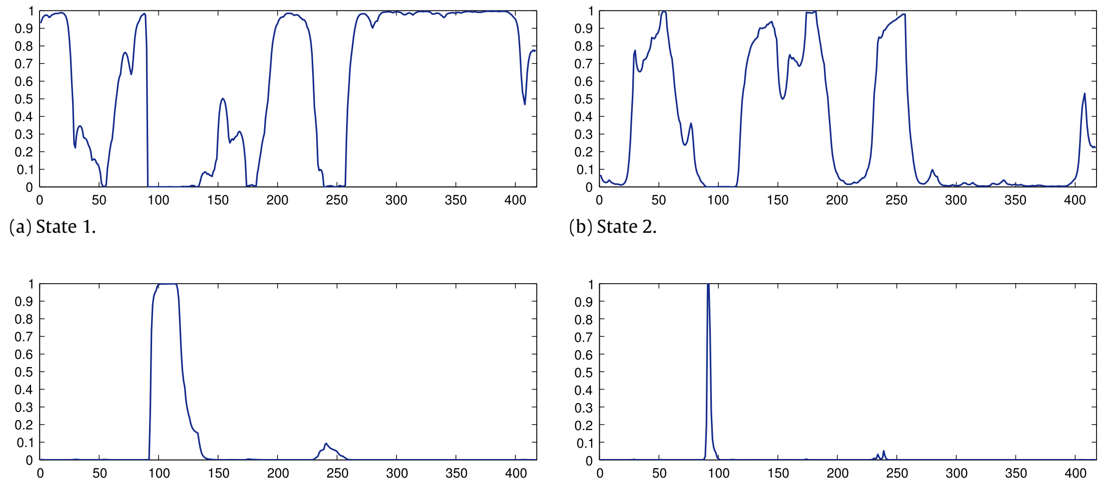
(a) State 1.

(b) State 2.

(c) State 3.

(d) State 4.

Fig. 3. Filtered probabilities in the four-states case.



(a) State 1.

(b) State 2.

(c) State 3.

(d) State 4.

Fig. 4. Smoothed probabilities in the four-states case.

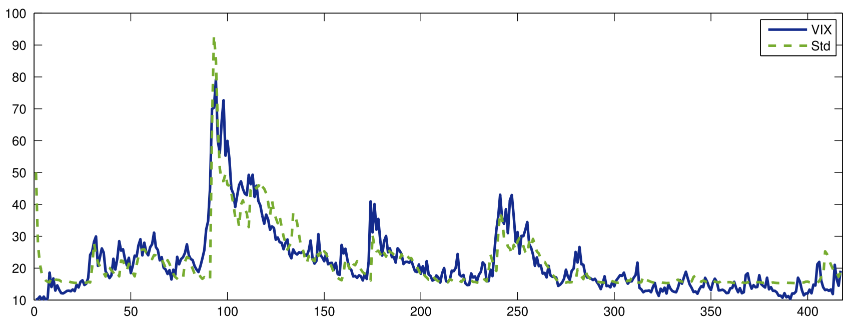


Fig. 5. Graphical comparison between standard deviation estimate and the VIX index. The estimate has been multiplied by a constant c in order to minimize the mean distance between the two graphs in the four-states case. Scale factor $c = 1322, 4$.

Another key feature we want to analyse is the global goodness of the regressed model. In particular we want to perform a *goodness of fit* analysis like the one proposed in Section 4. First we compute the standardized residuals $\hat{\epsilon}_t$. If the model is a good fit for our time series, the residuals will be generated by a Gaussian distribution. In Figs. 6–8 we report the related plots, the histogram and the *normal probability plot* (NPP) of the standardized residuals.

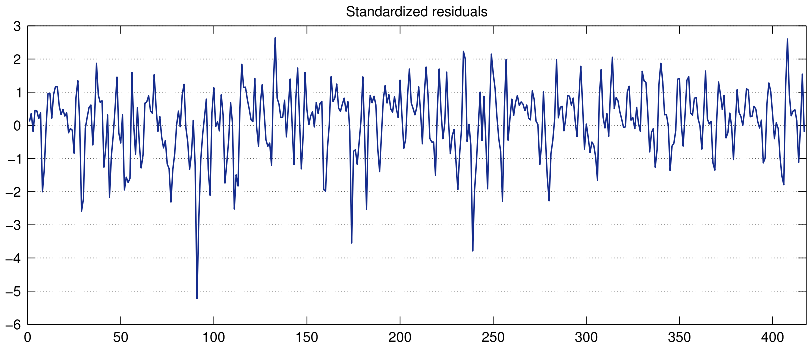


Fig. 6. Plot of the standardized residuals in the four-states case.

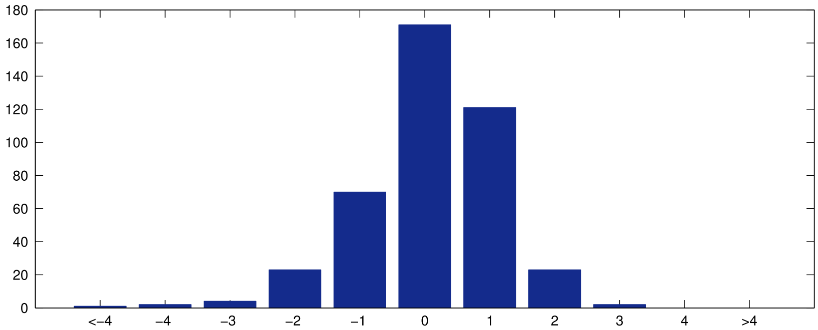


Fig. 7. Histogram of the standardized residuals in the four-states case.

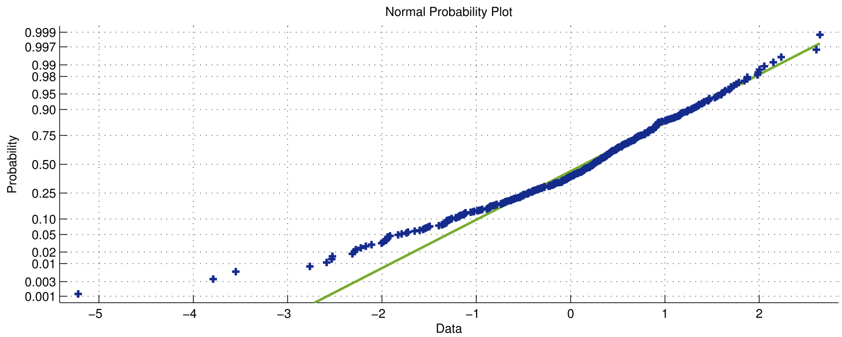


Fig. 8. Normal Probability Plot of the standardized residuals in the four-states case.

Applying two normality tests on $\{\hat{\epsilon}_t\}$, namely the Jarque–Bera test and the Lilliefors test, we can see that in both cases the null hypothesis of normality for the standardized residuals can be rejected at the 5% level. In order to pass the Jarque–Bera test, with 5% significance and p-value equal to 0.0642, we need to discard the first ten left outliers, which means 2.39% of the sample. Excluding those values restores the symmetry of the sample. We focused on the left tail because from the normal probability plot it is clear that the main outliers' contributions come from this side. On the Lilliefors test side we need to discard the first twenty-one left outliers to pass the test with 5% significance and p-value 0.0601, which means discarding 5.02% of the sample.

5.1.1. Comparison with three-states analysis

Further tests on three-state models highlighted the necessity of a four-state analysis for highly chaotic financial periods, which is exactly the case for the period we have taken into account. The regressed parameters are

$$P = \begin{bmatrix} 0.9783 & 0.0169 & 0.0048 \\ 0.0408 & 0.9592 & 0.0000 \\ 0.0000 & 0.0404 & 0.9596 \end{bmatrix},$$

$$\sigma_1^2 = 0.0001, \qquad \sigma_2^2 = 0.0004, \qquad \sigma_3^2 = 0.0021.$$

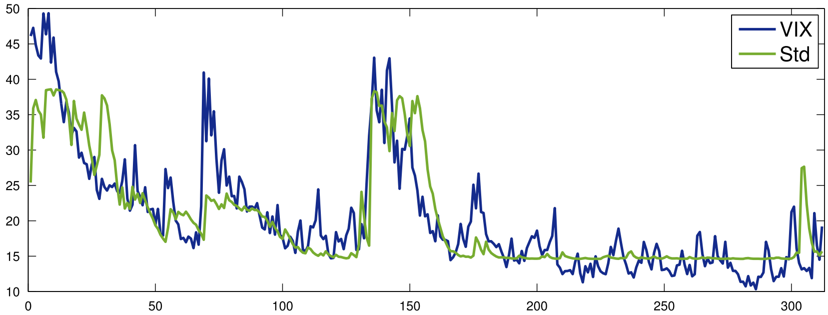


Fig. 9. Visual comparison between estimated volatility and VIX in the three-states case.

We would like to underline the relevant performance realized by the Markov Switching approach proposed in the previous chapters, where the high-volatility state has been included without influencing the measures on the low end the medium volatility regimes. Comparing the estimates for σ_1^2 and σ_2^2 in the three-states and in the four-states case we can see that the estimated values are the same, while the *high*-volatility state was split in an *a-bit-less-high*-volatility state and a *very high*-volatility state. The latter is a relevant result since, taking into consideration the estimated values, we see that the three-state model does not recognize the hidden risky nature of the financial asset that is crucial in the portfolio selection.

5.2. Short term: 2009–2014

In what follows we show the analysis performed focusing on the period between January 1st 2009 and December 31st 2014. Latter choice, namely to do not consider the first two years of the previously analysed time series, allows us to skip the great financial instability caused by the world financial crisis originated by the sub-prime bubble of the 2007–2008 biennium. The studied model is the same as the one of the long term period, and the associated four-state regression gives us the following results:

$$P = \begin{bmatrix} 0.9999 & 0.0000 & 0.0000 & 0.0000 \\ 0.0011 & 0.0010 & 0.0044 & 0.9934 \\ 0.0000 & 0.0109 & 0.9890 & 0.0000 \\ 0.0189 & 0.0002 & 0.0204 & 0.9605 \end{bmatrix},$$

$$\sigma_1^2 = 0.000113, \qquad \sigma_2^2 = 0.000001, \qquad \sigma_3^2 = 0.000219, \qquad \sigma_4^2 = 0.000887. \tag{30}$$

Observing the values for σ_S^2 we can check that they are not ordered, and this is an evidence factor for the hypothesis of fake regression. We reject the four-state MSM approach for the analysed time series.

Using a three-state approach we obtain the following results:

$$P = \begin{bmatrix} 0.9938 & 0.0000 & 0.0062 \\ 0.0205 & 0.9795 & 0.0000 \\ 0.0157 & 0.0265 & 0.9578 \end{bmatrix},$$

$$\sigma_1^2 = 0.000117, \qquad \sigma_2^2 = 0.000283, \qquad \sigma_3^2 = 0.000878. \tag{31}$$

Observing the graphical comparison between the estimated implied volatility and the VIX index (Fig. 9) we can notice, as we did in the long-term case, a smoother behaviour of the computed values with respect to the reference index and the same tendency to underestimate the peaks, but we can globally consider the estimated implied volatility a valid substitute for the VIX index.

In Fig. 10(c) it is possible to notice the presence of a very long left tail in the standardized residuals, we then expect the Jarque–Bera test and the Lilliefors test to fail at the 5% significance level, as is the case. In order to let the time series pass the normality tests, at the 5% significance level, we need to discard the three lower values, that means cutting 0.96% of the data, which is statistically irrelevant, for the Jarque–Bera test. Cutting the first nine values is necessary for the Lilliefors test to be passed.

6. Conclusion and future developments

We proved how Markov Switching Models can be successfully used to analyse financial time series, particularly when chaotic behaviours characterize their. The number of states that an operator should consider depends on his investment strategy: if he wants to maintain a *low*-volatility profile he can base his decisions on a three-states model, otherwise he needs to implement a four-states model, as we have shown in the present work. A further problem that should be faced is the one highlighted in the comparisons between the estimated implied volatility and the VIX, namely even if the changes of state

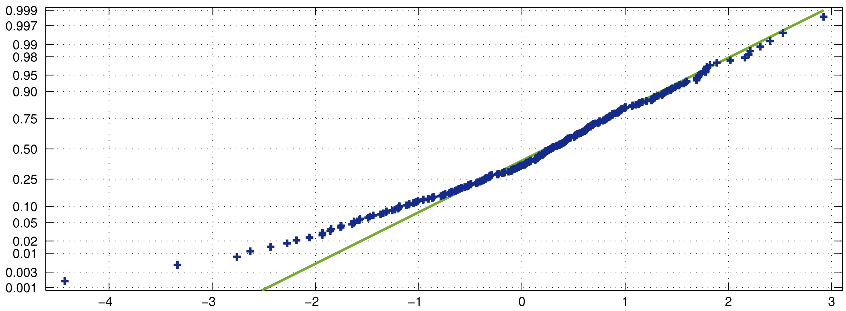


Fig. 10. Three-states case analysis of the standardized residuals.

and the peaks are correctly estimated, there is the necessity to improve the model in order to get a less smooth behaviour. In fact the latter result would allow for a more coherent analysis of breaks and peaks in relation with more complex volatility measures, as in the case of the VIX index. Another intriguing problem concerns the fact that the Gibbs Sampling approach needs an a-priori knowledge of the priors' hyperparameters, certainly a non trivial task, since such a knowledge requires extensive econometric studies which are rather expensive in terms of time as well as from a computational point of view.

References

[1] Romeo Negrea, Andrei Eckstein, Florin Alexa, Numerical solutions for a class of nonlinear systems and application to stochastic resonance, WSEAS Trans. Math. 9 (3) (2010) 161–170.
[2] Paulo Sérgio Ceretta, Ani Caroline Gigion Potrich, Pamela Amado, The relationship between oil and corn prices: Break-time versus regime-switching analysis, WSEAS Trans. Bus. Econ. 12 (2015).
[3] Mohammad Karimi, A convex optimization approach to control design for switching power converters with time delays, WSEAS Trans. Syst. Control 8 (2013).
[4] Luca Di Persio, Samuele Vettori, Markov switching model analysis of implied volatility for market indexes with applications to S&P 500 and DAX, J. Math. 2014 (2014).
[5] Chang-Jin Kim, Charles R. Nelson, State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches With Applications, vol. 2, MIT Press, Cambridge, 1999.
[6] Andrew Gelman, et al., Bayesian Data Analysis, in: Chapman & Hall/CRC Texts in Statistical Science, 2003.
[7] Daniel Sorensen, Daniel Gianola, Likelihood, Bayesian, and MCMC methods in Quantitative Genetics, Springer Science & Business Media, 2002.
[8] James H. Albert, Siddhartha Chib, Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, J. Bus. Econom. Statist. 11 (1) (1993) 1–15.
[9] Stuart Geman, Donald Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984) 721–741.
[10] Chris K. Carter, Robert Kohn, On Gibbs sampling for state space models, Biometrika 81 (3) (1994) 541–553.
[11] Luca Di Persio, Matteo Frigo, Maximum likelihood approach to Markov switching models, WSEAS Trans. Bus. Econ. 12 (2015) 239–242.
[12] Carlos M. Jarque, Anil K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Econom. Lett. 6 (3) (1980) 255–259.
[13] Hubert W. Lilliefors, On the Kolmogorov–Smirnov test for normality with mean and variance unknown, J. Amer. Statist. Assoc. 62 (318) (1967) 399–402.
[14] Thorsten Thadewald, Herbert Büning, Jarque–Bera test and its competitors for testing normality: a power comparison, J. Appl. Stat. 34 (1) (2007) 87–105.
[15] James Durbin, Geoffrey S. Watson, Testing for serial correlation in least squares regression. I, Biometrika 37 (3–4) (1950) 409–428.
[16] James Durbin, Geoffrey S. Watson, Testing for serial correlation in least squares regression. II, Biometrika (1951) 159–177.
[17] Gidon Eshel, The yule walker equations for the AR coefficients, Lect. Notes Stat. (2009) 956 course.