2016 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 13–16, 2016, SALERNO, ITALY

UNSUPERVISED LEARNING OF MARKOV-SWITCHING STOCHASTIC VOLATILITY WITH AN APPLICATION TO MARKET DATA

Ivan Gorynin, Emmanuel Monfrini and Wojciech Pieczynski

SAMOVAR, Telecom SudParis, CNRS, Universit Paris-Saclay, 9, rue Charles Fourrier, 91000 Evry, France.

ABSTRACT

We introduce a new method for estimating the regime-switching stochastic volatility models from the historical prices. Our methodology is based on a novel version of the assumed density filter (ADF). We estimate the switching model by maximizing the quasi-likelihood function of our ADF. The simulation experiments show the efficiency of our method. Then we analyze different market price histories for consistency with a regime-shifting model.

***Index Terms*—** Stochastic volatility, Markov-switching stochastic volatility models, Quasi-maximum likelihood, Assumed density filtering, Gaussian quadrature.

1. INTRODUCTION

In 1973 Black and Scholes [1] introduce the concept of volatility of asset price and propose a methodology for valuating the related derivatives. A derivative is a security that derives its value from the characteristics of the underlying entity. This entity can be an asset, index, exchange or interest rate or even some other derivative. Its characteristics include its value, yield and volatility. The global derivatives market is gigantic and is estimated at more than \$1.2 quadrillion.

The interest of Black's theory is that allows estimating a "fair" value of a derivative. This enables traders to buy underpriced securities as well as to sell overpriced ones and make profit on average. The theory assumes that the infinitesimal return of the underlying follows a Brownian motion, *i.e.*

$$\frac{\mathrm{d}S_t}{S_t} = \mu_t \mathrm{d}t + h_t \mathrm{dB}_t.$$

We denote the price of the underlying at the time instant t by $S_t \in \mathbb{R}$, B_t is a Brownian motion in $\mathbb{R}$, $\mu_t \in \mathbb{R}$ is a drift and $h_t \in \mathbb{R}_+$ is the volatility.

The volatility of an asset is a virtual concept and is not observable directly. The stochastic volatility (SV) models reproduce the joint dynamics of both the volatility and the returns of a given equity. The idea is to estimate these models from the historical prices available between a past and today's time instants t_0 and t. The papers [2, 3] are concerned with the techniques to perform this estimation.

The Markov-switching stochastic volatility models have a jump component in the mean of the volatility process. The estimation of such models is not trivial. There are several simulations-based approaches in the literature *e.g.* [4, 5].

The quasi-maximum likelihood (QML) methods [6, 7, 8] are alternative to the simulations-based ones. In fact, the likelihood function in the SV models is usually intractable. A QML approach consists in maximizing the log-likelihood function of an approximation of the true likelihood. These estimators are less computationally demanding than their simulations-based counterparts [8]. Besides, they do not introduce any hyperparameters and thus do not suffer from the sensitivity to them [4]. The classic QML estimators [6, 7] intrinsically assume an online linear (Kalman) relationship between the log-volatility and the log-squared returns. However, several researches motivate the need to avoid such a linearization and offer the linearization-free QML estimators [9, 8].

This work takes up the idea from [8] and extends it to the switching volatility models. To do so, we derive a linearization-free assumed density filter (ADF) [10] which handles the switches.

The object of the next section is to recall the autoregressive [11], asymmetric [7, 12] and Markov-switching [13, 14] stochastic volatility models. Third section contains a detailed derivation of our approach. Fourth section contains experiments on both synthetic and market data and fifth section is a conclusion.

2. STOCHASTIC VOLATILITY MODELS

2.1. Basic stochastic volatility models

The autoregressive stochastic volatility (ARSV) model is a primary SV model and is subject to numerous extensions. This model assumes that the logarithm of the variance h_t^2 follows an autoregressive process. Let Δt denote a constant period of time between two successive observations and n is a positive time index. Next, we denote by Y_n the log-return of the underlying at day n and X_n is the corresponding log-variance:

$$Y_n = \log \frac{S_{t_0+n\Delta t}}{S_{t_0+(n-1)\Delta t}}; \quad (1a)$$

$$X_n = 2\log h_{t_0+n\Delta t}. \quad (1b)$$

In this setting, the ARSV equations are

$$X_{n+1} = \alpha + \phi X_n + \sigma U_{n+1}; \quad (2a)$$

$$Y_n = \exp(X_n \backslash 2) V_n, \quad (2b)$$

where $\{U_n\}_{n\geq 1}$, $\{V_n\}_{n\geq 1}$ are standard Gaussian white noises and (α, ϕ, σ) are fixed parameters, $|\phi| \leq 1$.

The asymmetric stochastic volatility (ASV) model assumes that the innovations $\{U_n\}_{n\geq 1}$ in the log-variance process correlate with the disturbances $\{V_n\}_{n\geq 1}$. That is to reproduce the increase in volatility that follows a drop in the equity returns. This effect is known as leverage. The ASV equations are:

$$X_{n+1} = \alpha + \phi X_n + \sigma\rho \exp(-X_n \backslash 2) Y_n + \lambda U_{n+1}; \quad (3a)$$

$$Y_n = \exp(X_n \backslash 2) V_n, \quad (3b)$$

where $\{U_n\}_{n\geq 1}$ and $\{V_n\}_{n\geq 1}$ are standard Gaussian white noises, ρ is a fixed parameter such that $-1 < \rho \leq 0$ and $\lambda = \sigma\sqrt{1-\rho^2}$. This value of λ ensures that σ is the standard deviation of the innovations in the log-variance process when seen independently from $\{Y_n\}_{n\geq 1}$.

2.2. Markov-switching stochastic volatility models

The classic Markov-switching stochastic volatility model (MSSV) introduces a regime-shifting effect through a Markov chain $\{R_n\}_{n\geq 1}$. This chain has k possible states, *i.e.* for each positive n, $R_n \in \{1, .., k\}$. The transition probabilities of $\{R_n\}_{n\geq 1}$ are independent from n, and we note them as

$$p_{j|i} = P\left[R_{n+1} = j \,|R_n = i\right]. \quad (4)$$

Note that we do deliberately not consider the initial state probabilities $P\left[R_1 = i\right]$ for $1 \leq i \leq k$. In practice, a switching model is not sensitive to them due to the rapid mixing property of the Markov chains [15]. Therefore, we assume that the distribution of R_1 is invariant with respect to the Markov transition (4).

The MSSV model extends the ARSV by making α in (2) depend on the current state of the Markov chain. We write it down as

$$X_{n+1} = \alpha_{R_{n+1}} + \phi X_n + \sigma U_{n+1}; \quad (5a)$$

$$Y_n = \exp(X_n \backslash 2) V_n, \quad (5b)$$

with the same assumptions as for the ARSV model. The overall parameter is $\theta = \{\alpha_i, \phi, \sigma, p_{j|i} \,|1 \leq i, j \leq k\}$.

Similarly, the Markov-switching asymmetric stochastic volatility (MSASV) model is

$$X_{n+1} = \alpha_{R_{n+1}} + \phi X_n + \sigma\rho \exp(-X_n \backslash 2) Y_n + \lambda U_{n+1}; \quad (6a)$$

$$Y_n = \exp(X_n \backslash 2) V_n, \quad (6b)$$

with the same assumptions as for the ASV model.

Finally, let us present our Markov-switching asymmetric stochastic volatility surrogate (MSASVS) model. It covers a broad range of the switching SV models. We define its equations as follows:

$$X_{n+1} = f^{\theta}_{n+1}\left(X_n, R_{n+1}, Y_n\right) + \sigma^{\theta}_{n+1}\left(R_{n+1}, Y_n\right) U_{n+1}; \quad (7a)$$

$$Y_n = \exp(X_n \backslash 2) V_n, \quad (7b)$$

where $\{U_n\}_{n\geq 1}$ and $\{V_n\}_{n\geq 1}$ are standard Gaussian white noises. The notation f^{θ}_{n+1} , σ^{θ}_{n+1} relates to SV model-dependent functions which take random variables and the parameter vector θ as argument. We explain further how we estimate θ. However, these functions derive from a specific switching SV model so our algorithm does not construct them but assumes known. We introduce this meta-model to make wider the scope of application of our method.

3. PARAMETER INFERENCE WITH ASSUMED DENSITY FILTERING

The maximum-likelihood estimation (MLE) is a well-known method of estimating the parameters of a statistical model. It searches for the most likely value of the unknown parameter vector θ that would reproduce the observed time series $y_{1...N} = \{y_n\}_{n\geq 1}$. In other words, it maximizes

$$\theta \to p^{\theta}\left(y_{1...N}\right).$$

Regarding the models (5)-(7), the likelihood function is usually intractable. The QML approach consists in maximizing an approximation to the log-likelihood function of (7) (*i.e.* a quasi-likelihood function) instead of its exact likelihood. Next, for a quasi-likelihood function of $\{y_n\}_{n=1}^{N}$

$$\ell : \theta \to \ell\left(\theta, y_{1...N}\right), \quad (8)$$

the related parameter estimation is defined by

$$\widehat{\theta}(y_{1...N}) = \arg\max_{\theta} \ell\left(\theta, y_{1...N}\right). \quad (9)$$

Let us note that each method of evaluation of quasi-likelihood function is related to a QML estimator. In fact, any nonlinear optimization routine can solve (9). The processing time depends on the complexity of evaluation of (8).

It is common to derive a quasi-likelihood function by using the assumed density filtering (ADF). Below we derive a specific algorithm for the switching stochastic volatility models.

3.1. Quasi-likelihood evaluation in the switching stochastic volatility models

A variant of the ADF for the switching linear models is available in [16]. We generalize it to handle the MSASVS (7) dynamics. In what follows, p^θ denotes the true probability mass function and q^θ is its projection in the ADF workspace. By reproducing the reasoning [16] we consider, for each $n > 0$ and for each r_n in $\{1,..,k\}$, a normal approximation $q^\theta\left(x_n\,|r_n, y_{1\ldots n}\right)$ as follows:

$$q^\theta\left(x_n\,|r_n, y_{1\ldots n}\right) = \mathcal{N}\left(x_n; \widehat{x}_{n|n}(r_n), \widehat{\Gamma}_{n|n}(r_n)\right).$$

The ADF recursion computes $\widehat{x}_{n|n}(r_n)$ and $\widehat{\Gamma}_{n|n}(r_n)$ for any n with a linear complexity. Then we define, for each (r_n, r_{n+1}), a normal prediction mass

$$q^\theta\left(x_{n+1}\,|r_n, r_{n+1}, y_{1\ldots n}\right) = \mathcal{N}\left(x_{n+1}; \widehat{x}_{n+1|n}(r_n, r_{n+1}), \widehat{\Gamma}_{n+1|n}(r_n, r_{n+1})\right).$$

However, one supposes classically that the joint mass $q^\theta\left(x_{n+1}, y_{n+1}\,|r_n, r_{n+1}, y_{1\ldots n}\right)$ is normal too. This results in a linear (Kalman) relationship between x_{n+1} and y_{n+1}. In order to avoid this, we assume that only the conditional mass $q^\theta\left(x_{n+1}\,|r_n, r_{n+1}, y_{1\ldots n+1}\right)$ is normal. We have

$$q^\theta\left(x_{n+1}\,|r_n, r_{n+1}, y_{1\ldots n+1}\right) = \mathcal{N}\left(x_{n+1}; \widehat{x}_{n+1|n+1}(r_n, r_{n+1}), \widehat{\Gamma}_{n+1|n+1}(r_n, r_{n+1})\right).$$

We make some shortcuts to reduce the amount of notation:

$$\widehat{r}_{n|n}(i) = q^\theta\left(r_n = i\,|y_{1\ldots n}\right);$$
$$\widehat{r}_{n+1|n}(i,j) = q^\theta\left(r_n = i, r_{n+1} = j\,|y_{1\ldots n}\right);$$
$$\widehat{r}_{n+1|n+1}(i,j) = q^\theta\left(r_n = i, r_{n+1} = j\,|y_{1\ldots n+1}\right).$$

At each filter step, we begin classically by computing $\widehat{x}_{n+1|n+1}(r_n, r_{n+1})$ and $\widehat{\Gamma}_{n+1|n+1}(r_n, r_{n+1})$ for each (r_n, r_{n+1}) in $\{1,..,k\}^2$. Then we find $\widehat{x}_{n+1|n+1}(r_{n+1})$ and $\widehat{\Gamma}_{n+1|n+1}(r_{n+1})$ for each r_{n+1} in $\{1,..,k\}$ by matching the first two moments. We recall that these moments suffice to define a normal distribution.

Algorithm 1 ***Quasi-likelihood evaluation for a regime-switching stochastic volatility model***
Input: *Functions f^θ_{n+1}, σ^θ_{n+1} for positive n, parameter vector θ with respect to* (7) *and observed time series $\{y_n\}_{n=1}^N$.*
Output: *The quasi-log likelihood $\mathcal{L} = \log \ell\left(\theta, y_{1\ldots N}\right)$*
Initialization:

(i) For each (r_0, r_1) in $\{1,..,k\}^2$, assign to $\widehat{x}_{1|0}(r_0, r_1)$, $\widehat{\Gamma}_{1|0}(r_0, r_1)$ the prior mean and variance of X_1;

(ii) For each (i,j) in $\{1,..,k\}^2$, $\widehat{r}_{1|0}(i,j) \leftarrow p^\theta\left(r_1 = j\right)$;

(iii) Assign zero to $\mathcal{L}$;

Recursion: *for n in $\{0,\ldots,N-1\}$, repeat:*

(A) For each (i,j) in $\{1,..,k\}^2$:

(a) Let $\omega(x_n) = \mathcal{N}\left(x_n; \widehat{x}_{n|n}(r_n), \widehat{\Gamma}_{n|n}(r_n)\right)$;

(b) Assign to $\widehat{x}_{n+1|n}(i,j)$ and $\widehat{\Gamma}_{n+1|n}(i,j)$ respectively

$$\widehat{x}_{n+1|n}(i,j) = \int f^\theta_{n+1}\left(x_n, j, y_n\right)\omega(x_n)\mathrm{d}x_n; \quad \text{(10a)}$$

$$\widehat{\Gamma}_{n+1|n}(i,j) = \left(\sigma^\theta_{n+1}\left(j, y_n\right)\right)^2 + \int \left(f^\theta_{n+1}\left(x_n, j, y_n\right) - \widehat{x}_{n+1|n}(i,j)\right)^2 \omega(x_n)\mathrm{d}x_n; \quad \text{(10b)}$$

(c) Let $\omega'\left(x_{n+1}\right) = \mathcal{N}\left(x_{n+1}; \widehat{x}_{n+1|n}(i,j), \widehat{\Gamma}_{n+1|n}(i,j)\right)$;

(d) Assign to $c_{n+1}\left(\theta, i, j, y_{1\ldots n+1}\right)$

$$\int \mathcal{N}\left(y_{n+1}; 0, \exp x_{n+1}\right)\omega'(x_{n+1})\mathrm{d}x_{n+1}; \quad \text{(11)}$$

(e) Define $\omega''\left(x_{n+1}\right)$ by

$$\omega''(x_{n+1}) = \frac{\omega'(x_{n+1})\mathcal{N}\left(y_{n+1}; 0, \exp x_{n+1}\right)}{c_{n+1}\left(\theta, i, j, y_{1\ldots n+1}\right)};$$

(f) Assign to $\widehat{x}_{n+1|n+1}(i,j)$ and $\widehat{\Gamma}_{n+1|n+1}(i,j)$ respectively

$$\int x_{n+1}\omega''(x_{n+1})\mathrm{d}x_{n+1}; \quad \text{(12a)}$$

$$\int \left(x_{n+1} - \widehat{x}_{n+1|n+1}(i,j)\right)^2 \omega''(x_{n+1})\mathrm{d}x_{n+1}; \quad \text{(12b)}$$

(g) $\widehat{r}_{n+1|n}(i,j) = \widehat{r}_{n|n}(i)p_{j|i}$.

(B) Assign to $c_{n+1}\left(\theta, y_{1\ldots n+1}\right)$

$$\sum_{1\le i,j\le k} \widehat{r}_{n+1|n}(i,j)c_{n+1}\left(\theta, i, j, y_{1\ldots n+1}\right);$$

(C) For each (i,j) in $\{1,..,k\}^2$,

$$\widehat{r}_{n+1|n+1}(i,j) = \frac{\widehat{r}_{n+1|n}(i,j)c_{n+1}\left(\theta, i, j, y_{1\ldots n+1}\right)}{c_{n+1}\left(\theta, y_{1\ldots n+1}\right)};$$

(D) For each j in $\{1,..,k\}$,

$$\widehat{r}_{n+1|n+1}(j) = \sum_{1\le i\le k} \widehat{r}_{n+1|n+1}(i,j);$$

(E) Let $\forall(i,j) \in \{1,..,k\}^2$,

$$\beta(i,j) = \frac{\widehat{r}_{n+1|n+1}(i,j)}{\widehat{r}_{n+1|n+1}(j)};$$

(F) For each j in $\{1,..,k\}$:

$$\widehat{x}_{n+1|n+1}(j) = \sum_{i=1}^{k} \beta(i,j)\widehat{x}_{n+1|n+1}(i,j);$$

$$\widehat{\Gamma}_{n+1|n+1}(j) = \sum_{i=1}^{k} \beta(i,j)\widehat{\Gamma}_{n+1|n+1}(i,j) \quad + \sum_{i=1}^{k} \beta(i,j)\left(\widehat{x}_{n+1|n+1}(j) - \widehat{x}_{n+1|n+1}(i,j)\right)^2;$$

(G) Assign $\mathscr{L} \leftarrow \mathscr{L} + \log c_{n+1}(\theta, y_{1\dots n+1})$.

Steps (B) - (G) are well known and are the same for both linear and non-linear switching models. Steps (a) - (b) are also standard and should be ignored at the first iteration (at $n = 0$). Steps (d),(f) replace the classic Kalman linear update. Instead, we compute the involved integrals by using the Gaussian quadrature. This cancels the linearization-induced bias [8, 9, 10].

3.2. Gaussian quadrature

We see that **Algorithm** 1 includes evaluating of the integrals in (10)-(12). These integrals are of the form

$$I = \int g(\mathbf{x})\mathscr{N}\left(\mathbf{x}; \mathbf{m}, \mathbf{s}^2\right) \mathrm{d}\mathbf{x}, \tag{13}$$

where g is a function, $\mathbf{m}$ and $\mathbf{s}^2$ are the mean and the variance of the normal mass function in the integral. The value of I also is

$$I = \int g(\mathbf{m} + \mathbf{s}\mathbf{x})\mathscr{N}(\mathbf{x}; 0, 1) \mathrm{d}\mathbf{x}. \tag{14}$$

The Gaussian quadrature is a computational approach which allows approximating the value of I in (14) by:

$$I \approx \sum_{\boldsymbol{q}=1}^{M} \pi_{\boldsymbol{q}} g(\mathbf{m} + \mathbf{s}\xi_{\boldsymbol{q}}). \tag{15}$$

where $\{\xi_{\boldsymbol{q}}\}_{\boldsymbol{q}=1}^{M}$ are M points in $\mathbb{R}$ (known as the quadrature nodes) and $\{\pi_{\boldsymbol{q}}\}_{\boldsymbol{q}=1}^{M}$ are their corresponding weights. Such an approximation is exact if g is a polynomial function up to the $(2M-1)^{\text{th}}$ order *cf.* [17].

To find these nodes and weights, consider a symmetric tridiagonal matrix $\boldsymbol{J}$ of size $M \times M$ such that its diagonal elements are null and $\boldsymbol{J}[m, m+1] = \sqrt{\frac{m-1}{2}}, m \in \{1,..,M-1\}$. Next, find the eigenvalues $\{\epsilon_{\boldsymbol{q}}\}_{\boldsymbol{q}=1}^{M}$ and the normalized eigenvectors $\{\boldsymbol{v}_{\boldsymbol{q}}\}_{\boldsymbol{q}=1}^{M}$ of $\boldsymbol{J}$. Then the quadrature nodes and weights are:

$$\xi_{\boldsymbol{q}} = \sqrt{2}\epsilon_{\boldsymbol{q}}, \pi_{\boldsymbol{q}} = \left([\boldsymbol{v}_{\boldsymbol{q}}]_1\right)^2, \tag{16a}$$

where $[\boldsymbol{v}_{\boldsymbol{q}}]_1$ stands for the first element of $\boldsymbol{v}_{\boldsymbol{q}}$. The formulas below carry out a numerical integration involved by steps (a)-(f) of **Algorithm** 1.

(a) Let $\forall \boldsymbol{q} \in \{1, \dots, M\}$,

$$\xi'_{\boldsymbol{q}} = \widehat{x}_{n|n}(i) + \sqrt{\widehat{\Gamma}_{n|n}(i)}\xi_{\boldsymbol{q}};$$

(b) Assign to $\widehat{x}_{n+1|n}(i,j)$ and $\widehat{\Gamma}_{n+1|n}(i,j)$ respectively

$$\widehat{x}_{n+1|n}(i,j) = \sum_{\boldsymbol{q}=1}^{M} \pi_{\boldsymbol{q}} f_{n+1}^{\theta}\left(\xi'_{\boldsymbol{q}}, j, y_n\right);$$

$$\widehat{\Gamma}_{n+1|n}(i,j) = \left(\sigma_{n+1}^{\theta}(j, y_n)\right)^2 \quad + \sum_{\boldsymbol{q}=1}^{M} \pi_{\boldsymbol{q}} \left(f_{n+1}^{\theta}\left(\xi'_{\boldsymbol{q}}, j, y_n\right) - \widehat{x}_{n+1|n}(i,j)\right)^2;$$

(c) Let $\forall \boldsymbol{q} \in \{1, \dots, M\}$,

$$\xi''_{\boldsymbol{q}} = \widehat{x}_{n+1|n}(i,j) + \sqrt{\widehat{\Gamma}_{n+1|n}(i,j)}\xi_{\boldsymbol{q}};$$

(d) Assign to $c_{n+1}(\theta, i, j, y_{1\dots n+1})$

$$\sum_{\boldsymbol{q}=1}^{M} \pi_{\boldsymbol{q}} \mathscr{N}\left(y_{n+1}; 0, \exp \xi''_{\boldsymbol{q}}\right);$$

(e) Let $\forall \boldsymbol{q} \in \{1, \dots, M\}$,

$$\pi''_{\boldsymbol{q}} = \frac{\pi_{\boldsymbol{q}} \mathscr{N}\left(y_{n+1}; 0, \exp \xi''_{\boldsymbol{q}}\right)}{c_{n+1}(\theta, i, j, y_{1\dots n+1})};$$

(f) Assign to $\widehat{x}_{n+1|n+1}(i,j)$ and $\widehat{\Gamma}_{n+1|n+1}(i,j)$ respectively

$$\widehat{x}_{n+1|n+1}(i,j) = \sum_{\boldsymbol{q}=1}^{M} \pi''_{\boldsymbol{q}} \xi''_{\boldsymbol{q}}; \tag{18a}$$

$$\sum_{\boldsymbol{q}=1}^{M} \pi''_{\boldsymbol{q}} \left(\xi''_{\boldsymbol{q}} - \widehat{x}_{n+1|n+1}(i,j)\right)^2; \tag{18b}$$

Finally, we obtain a QML parameter estimate $\widehat{\theta}$ by maximizing the output of **Algorithm** 1 with respect to θ. We use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization routine.

Remark 1 *The approximation error in* (15) *decreases at a rate of $\mathscr{O}\left(\frac{1}{M^2}\right)$* cf. *[18]. The corresponding Monte Carlo integration error decreases at $\mathscr{O}\left(\frac{1}{\sqrt{M'}}\right)$ where M' is the number of random integration nodes. The one-dimensional Gaussian quadrature converges faster. Compared to a simulation-based estimation, the proposed QML approach allows a better trade-off between the accuracy and the computation time. Note that the Gaussian quadrature can be combined with the sparse grid theory to compete the Monte-Carlo integration even in higher dimensions* cf. *[19].*

4. EXPERIMENTS

4.1. Simulation study

Let us consider the MSSV model (5) which has $k = 2$ regimes. We use the following parameters (*cf.* [4]): $\alpha_1 = -5$, $\alpha_2 = -2$, $\sigma = 0.32$, $\phi = 0.5$, $p_{1|1} = 0.990$, $p_{2|2} = 0.985$ for Test 1, then $p_{1|1} = 0.85$, $p_{2|2} = 0.25$ for Test 2 and $p_{1|1} = 0.5$, $p_{2|2} = 0.5$ for Test 3. In each test, the prior distribution of jumps corresponds to the eigenvector of the Markov transition matrix.

Let $N = 1000$, we perform the following experiment 100 times per each test:

- Simulate $\{x_n, y_n\}_{n=1}^N$ by using (5) with $\theta_0 = (p_{1|1}, p_{2|2}, \alpha_1, \alpha_2, \phi, \sigma)$ depending on the test;
- Perform a QML estimation by maximizing the likelihood function $\ell(\theta, y_{1...N})$ from **Algorithm** 1. We use $M = 5$ integration nodes in the quadrature rules;
- Run a particle filter with 2000 particles to find the supervised $\{\widehat{x}_n^{\mathbf{sup}}\}_{n=1}^N$ and unsupervised $\{\widehat{x}_n^{\mathbf{unsup}}\}_{n=1}^N$ log-variance estimates as follows:

$$\widehat{x}_n^{\mathbf{sup}} = \mathbb{E}_{\theta_0}[X_n | y_{1...n}], \widehat{x}_n^{\mathbf{unsup}} = \mathbb{E}_{\widehat{\theta}}[X_n | y_{1...n}];$$

- Compute the mean square errors (MSE) by

$$\text{MSE}^{\mathbf{sup}} = \frac{1}{N}\sum_{n=1}^{N}(\widehat{x}_n^{\mathbf{sup}} - x_n)^2;$$

$$\text{MSE}^{\mathbf{unsup}} = \frac{1}{N}\sum_{n=1}^{N}(\widehat{x}_n^{\mathbf{unsup}} - x_n)^2.$$

We report in Table 1 the average over these experiments and the average parameter estimates per each test.

This simulation study shows that the accuracy of our unsupervised method is satisfactory. We also observed that using a larger number of integration nodes does not affect the outcome.

4.2. Stock market volatility analysis

We have collected 100 historical prices of the most active stocks by trade volume, of the most volatile stocks and those of the largest companies by revenue.

These charts are publicly available via Yahoo! Finance service. In our analysis, they relate to the period in-between 01/01/2010 and 04/01/2016 and contain 1500 price observations each on average.

We estimate the parameters of bi-regime MSSV (5) and MSASV (6) models for each of these shares with our QML estimator. By analyzing these parameters we can find out whether or not the switching models are relevant for the of the price movement modeling. For example, if the parameters of

Table 1. Error measures for both supervised and unsupervised log-variance estimates in the MSSV model, as well as the true parameters and their average estimates. The percentage quantifies the increase in error when the parameter vector is unknown.

	Test 1	Test 2	Test 3
$\text{MSE}^{\mathbf{sup}}$	0.34	0.82	1.46
$\text{MSE}^{\mathbf{unsup}}$	0.37	0.83	1.47
Percentage	9.4%	1.9%	0.7%
$p_{1\|1}, p_{2\|2}$	$0.990, 0.985$	$0.85, 0.25$	$0.50, 0.50$
$\widehat{p}_{1\|1}, \widehat{p}_{2\|2}$	$0.991, 0.984$	$0.86, 0.26$	$0.49, 0.51$
α_1, α_2	$-5.0, -2.0$	$-5.0, -2.0$	$-5.0, -2.0$
$\widehat{\alpha}_1, \widehat{\alpha}_2$	$-5.5, -2.2$	$-5.1, -2.1$	$-5.3, -2.2$
ϕ, σ	$0.50, 0.32$	$0.50, 0.32$	$0.50, 0.32$
$\widehat{\phi}, \widehat{\sigma}$	$0.45, 0.19$	$0.48, 0.30$	$0.47, 0.30$

a MSSV are $p_{1|1} = 1$, $p_{2|2} = 0$, then the second regime is clearly absent and we can say that MSSV includes only one regime. The same holds if $\alpha_1 = \alpha_2$ *i.e.* the two regimes are exactly the same.

We find out that MSSV and MSASV have only one regime in 32 of the analyzed charts. We conclude that the related equities do not seem to have a regime-switching volatility. The MSSV and MSASV are identical for 36 equities from our list. In other words, the corresponding log-variance is not asymmetric but includes regime-shifting. The MSSV and MSASV produce two nonidentical regimes in 24 cases. The related shares have an asymmetric switched log-variance. Consider **Fig.** 1 for an illustration. There are only three shares for which the MSASV identifies two different regimes while MSSV identifies only one. The corresponding regimes are undetectable unless we use an asymmetric model. These shares concern large companies, specifically Apple, AO Smith and IBM. Finally, there are five charts where MSSV identifies two different regimes while MSASV identifies only one. Indeed, these charts include the well-known SPX and FTSE indexes. The corresponding log-variance is highly asymmetric so the asymmetric volatility model fits it better than non-asymmetric Markov-switching one.

5. CONCLUSION AND PERSPECTIVES

We proposed a general method of parameter estimation in the switched stochastic volatility models. The solution results from maximizing a specific approximation of the likelihood function. This method is simple, efficient and allows performing an analysis of stock quotes dynamics. Our approximation is designed to realize considerable speedups compared to the existing simulation-based techniques. A Monte-Carlo study confirms the effectiveness of our methodology.

The proposed method can be generalized to a broad range of common activities such as option pricing and yield curve

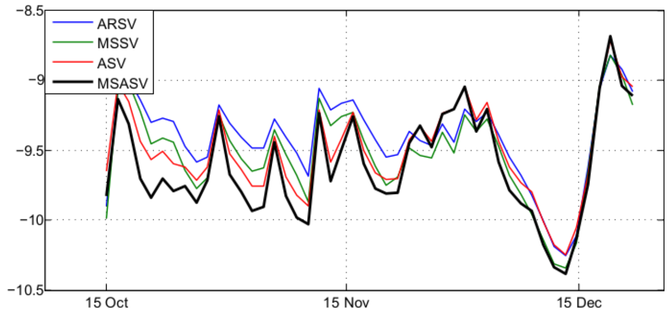

Fig. 1. Log-variance of the XLF index by the four models in-between 01/10/2015 and 08/12/2015 (48 observations).

fitting. As a perspective, we will study the applicability of this method to more complex models. Such are, for example, multi-factor, multi-scale volatility models. We will also propose tests for model selection.

6. REFERENCES

[1] Fischer Black and Myron Scholes, "The pricing of options and corporate liabilities," *The journal of political economy*, pp. 637–654, 1973.

[2] Sangjoon Kim, Neil Shephard, and Siddhartha Chib, "Stochastic volatility: likelihood inference and comparison with arch models," *The Review of Economic Studies*, vol. 65, no. 3, pp. 361–393, 1998.

[3] Eric Jacquier, Nicholas G Polson, and Peter E Rossi, "Bayesian analysis of stochastic volatility models," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 69–87, 2002.

[4] Yun Bao, Carl Chiarella, and Boda Kang, "Particle filters for markov switching stochastic volatility models," *Research Paper*, vol. 1, no. 299, pp. 1–21, 2012.

[5] Mai Shibata and Toshiaki Watanabe, "Bayesian analysis of a markov switching stochastic volatility model," *Journal of the Japan Statistical Society*, vol. 35, no. 2, pp. 205–219, 2005.

[6] Esther Ruiz, "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of econometrics*, vol. 63, no. 1, pp. 289–306, 1994.

[7] Andrew C Harvey and Neil Shephard, "Estimation of an asymmetric stochastic volatility model for asset returns," *Journal of Business & Economic Statistics*, vol. 14, no. 4, pp. 429–434, 1996.

[8] Hiroyuki Kawakatsu, "Numerical integration-based gaussian mixture filters for maximum likelihood estimation of asymmetric stochastic volatility models," *The Econometrics Journal*, vol. 10, no. 2, pp. 342–358, 2007.

[9] Onno Zoeter, Alexander Ypma, and Tom Heskes, "Improved unscented kalman smoothing for stock volatility estimation," in *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop*. IEEE, 2004, pp. 143–152.

[10] Thomas P Minka, *A family of algorithms for approximate Bayesian inference*, Ph.D. thesis, Massachusetts Institute of Technology, 2001.

[11] Stephen J Taylor, "Modeling stochastic volatility: A review and comparative study," *Mathematical finance*, vol. 4, no. 2, pp. 183–204, 1994.

[12] Yasuhiro Omori and Toshiaki Watanabe, "Block sampler and posterior mode estimation for asymmetric stochastic volatility models," *Computational Statistics & Data Analysis*, vol. 52, no. 6, pp. 2892–2910, 2008.

[13] Madhu Kalimipalli and Raul Susmel, "Regime-switching stochastic volatility and short-term interest rates," *Journal of Empirical Finance*, vol. 11, no. 3, pp. 309–329, 2004.

[14] Tsunehiro Ishihara and Yasuhiro Omori, "Markov switching asymmetric stochastic volatility model with application to TOPIX data -A Permutation Sampler Approach-," Tech. Rep. CARF-J-045, 2008.

[15] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer, *Markov chains and mixing times*, American Mathematical Soc., 2009.

[16] David Barber, "Expectation correction for smoothed inference in switching linear dynamical systems," *The Journal of Machine Learning Research*, vol. 7, pp. 2515–2540, 2006.

[17] Ienkaran Arasaratnam, Simon Haykin, and Robert J Elliott, "Discrete-time nonlinear filtering algorithms using gauss–hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[18] W Barrett, "Convergence properties of gaussian quadrature formulae," *The Computer Journal*, vol. 3, no. 4, pp. 272–277, 1961.

[19] Ivan Gorynin, Emmanuel Monfrini, and Wojciech Pieczynski, "Fast filtering with new sparse transition markov chains," *Statistical Signal Processing, 2016. SSP16. IEEE/SP 19th Workshop*, 2016 (accepted).