Gibbs Sampling Approach to Markov Switching Models in Finance

LUCA DI PERSIO
University of Verona
Department of Computer Science
Strada Le Grazie 15, 37100 Verona
ITALY
luca.dipersio@univr.it

VUKASIN JOVIC
University of Verona
Department of Computer Science
Strada Le Grazie 15, 37100 Verona
ITALY
jovic_vukasin@yahoo.com

Abstract: In the present paper we apply the *Gibbs Sampling* approach to estimate the parameters of a Markov Switching Model which we use to model financial time series. In particular, we estimate the standard deviation of the time series in order to obtain an indicator similar to the VIX index. The Markov Switching technique has been chosen because of the presence of exogenous factors which can have a large impact on the market, making it behave differently in different time periods. We also perform a case study on the S&P500 index for the period 3 January, 2007 - 29 December, 2014.

Key–Words: Markov Switching, α-stable Distribution, Gibbs Sampling, Finance, Financial time series

1 Introduction

Most standard approaches to time series analysis assume stationarity, i.e. that the mean and variance of the set of observations do not change over time. While this assumption is true for a large number of situations, it fails when the data being analyzed are prone to structural breaks.

The main idea behind Markov Switching Models (MSM) is that in order to catch these changes we allow the distribution of the observations to change over time by making it dependent on past observations as well as on the state. This model is of the form:

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

where $\psi_t := \{y_k : k = 1, \dots, t\}$, $\tilde{S}_t := \{S_1, ..., S_t\}$, $\Lambda = \{1, ..., M\}$ is the set of all states and g is the function that governs the transitions between the states. The function f defines how the observation at time t depends on $S_t, \theta,$ and ψ_{t-1}.

This type model is extremely useful for modeling financial or economic data. In econometrics two state models are the most prevalent. One of the states is used to model sluggish economic growth while the other is used for rapid expansion, see e.g. [8]. In finance three state models are the preference with the states being interpreted as low, medium and high volatility. We opt for a four state model in which the fourth state will be used for situations where there is very high volatility, caused by stock market crashes, economic crises etc.

One thing we would like to point out here is how the transition law g is modeled. In the simpler two state case the transition law tends to be more complicated that in the three state case. This can be considered as a kind of trade-off between a larger number of states and a more intricate transition law. In the analysis that follows we use Markov transition probabilities, meaning that the state at time $t+1$ depends only on the state at time t.

2 The Markov Switching Model

In the proposed model we assume that our data follow a symmetric α-stable distribution, more precisely $y_t \sim \mathcal{S}_{\alpha,0}(\gamma_{S_t}, \mu_{S_t})$. Here $\mathcal{S}_{\alpha,\beta}(\gamma, \mu)$ is the notation for an α-stable distribution with stability parameter α, skewness parameter β, scale parameter γ and location parameter μ. The full model is:

$$\begin{cases} y_t \sim \mathcal{S}_{\alpha,0}(\gamma_{S_t}, \mu_{S_t}) \\ S_t \in \{1, ..., M\} \\ \gamma_{S_t} = \gamma_j \text{ if } S_t = j, \ \forall j \in \{1, ..., M\} \\ \mu_{S_t} = \mu_j \text{ if } S_t = j, \ \forall j \in \{1, ..., M\} \\ \alpha \in (1, 2) \\ p_{ij} = \mathbb{P}(S_t = j | S_{t-1} = i) \\ \pi_0 = [\mathbb{P}(S_0 = 1), ..., \mathbb{P}(S_0 = M)] \end{cases} \tag{2}$$

The motivation for using the above model is twofold. First, financial data exhibit fat tails which can not be well described using the Normal distribution. We believe that the use of α-stable distributions with $\alpha < 2$ is much better suited to this problem. Second, financial data often exhibit structural breaks because of abrupt changes in the market. The sub-prime mortgage credit crisis of 2008 is a prime example of this which is why we make the scale and location parameters state-dependent.

Computationally this framework is not very convenient since (in general) there is no closed form for the density of an α-stable distribution. We circumvent this using the fact that y_t can be (conditionally) represented as a Normal random variable, see [3, 7]. This is done by introducing the random variable λ.

$$\begin{aligned} &\text{If:} && \lambda \sim \mathcal{S}_{\frac{\alpha}{2},1}(2\left(\cos(\frac{\pi\alpha}{4})\right)^{\frac{2}{\alpha}}, 0) \\ &\text{Then:} && y_t|\lambda \sim \mathcal{N}(\mu_{S_t}, \lambda\gamma^2_{S_t}) \end{aligned}$$

There are three main points that need to be addressed in order for MSM to be effective. First, the number of states in the model has to be specified or inferred from the data. Second, the parameter values and the transition probabilities need to be estimated. And finally, the state vector $\tilde{S}_T$ needs to be sampled somehow. In this paper we build on the four-state model proposed in [1], therefore there is no need to worry about the first point. We simply take $\Lambda = \{1, 2, 3, 4\}$. The four states will be interpreted as low, medium, high and very high volatility. We deal with the second point using Gibbs sampling. Finally, the problem in the third point is solved using the the Hamilton filter (see [4]) and a simulation method which can be found in [5].

3 Bayesian Inference

Bayesian Inference is a branch of statistical inference that assumes the parameter (or parameters) of a probability distribution to be randomly distributed according to a "prior" distribution. The Bayes' rule (together with observed data) is then used to generate the "posterior" distribution of the parameter (parameters). The posterior distribution can be interpreted as the distribution of the parameter once we have taken into account both our subjective belief about the parameter (the prior) and the data. Mathematically we can represent this model in the following way:

$$\begin{aligned} \theta &\sim \pi(\theta) \\ y|\theta &\sim f(y|\theta) \\ f(\theta|y) &= \frac{\pi(\theta)f(y|\theta)}{f(y)} \end{aligned} \tag{3}$$

Here $\pi(\theta)$ is the prior distribution, $f(y|\theta)$ is the distribution of the data (which depends on the parameter θ) and $f(\theta|y)$ is the posterior of θ. Finally, $f(y)$ is the marginal distribution of y, i.e.

$$f(y) = \int f(y,\theta)d\theta = \int \pi(\theta)f(y|\theta)d\theta \tag{4}$$

Clearly the choice of prior can have a large impact on the posterior. A particularly convenient form of prior is what is known as a *conjugate prior*. We say that a prior distribution is conjugate if the posterior distribution that is obtained from it belongs to the same family. A good example of this the Beta-Bernoulli pair.

$$\begin{aligned} \pi(\theta) &= \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ f(y|\theta) &= \binom{n}{y}\theta^y(1-\theta)^{n-y} \\ f(y|\theta) &\propto \theta^{(\alpha+y)-1}(1-\theta)^{(\beta+n-y)-1} \\ &\propto \text{Beta}(\alpha+y, \beta+n-y) \end{aligned}$$

It should be noted that it is not necessary for the prior to be conjugate. Picking a non-conjugate prior that results in a well known posterior is just as effective for what is needed in the next point, which is samples from the posterior distribution. In order to make this process as efficient as possible, a posterior which can be simulated using an inbuilt R package is desirable.

4 Gibbs Sampling

Assume that we have a model with k parameters, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ and we want to find the full posterior distribution $f(\theta_1, ..., \theta_k|\boldsymbol{y})$. This can be quite difficult since the multivariate simulation of distributions is much more tasking than its univariate counterpart. Gibbs sampling allows us to sample $f(\theta_1, ..., \theta_k|\boldsymbol{y})$ knowing only $f(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_k, \boldsymbol{y}), \forall i \in \{1, ..., k\}$.

4.1 Gibbs Sampler

Let N be the number of simulations we want to perform. We assign arbitrary starting values $(\theta^0_1, ..., \theta^0_k)$ to each of the parameters. Then, for every $j \in \{1, ..., N\}$, we do the following:

Step 1: Draw θ^j_1 from $f(\theta^j_1|\theta^{j-1}_2, ..., \theta^{j-1}_k, \boldsymbol{y})$.

Step 2: Draw θ^j_2 from $f(\theta^j_2|\theta^j_1, \theta^{j-1}_3, ..., \theta^{j-1}_k, \boldsymbol{y})$.

$\vdots$

Step k: Draw θ^j_k from $f(\theta^j_k|\theta^j_1, ..., \theta^j_{k-1}, \boldsymbol{y})$.

Using this method we can simulate each of the parameters in this model. The first J simulations are discarded (this is known as the *burn in period*) in order to get rid of the simulations that are a result of the arbitrary starting point $(\theta_1^0, ..., \theta_k^0)$. The remaining $N - J$ values are then assumed to be an approximation of the real distribution. The number of iterations should be chosen carefully. Too large an N can make the computations cumbersome while a small value might not allow the sampler to converge.

5 State Simulation

In this section our goal is to simulate the state vector $\tilde{S}_T$. In order to accomplish this we need to obtain the values $\mathbb{P}(S_1|\tilde{y}_1), ..., \mathbb{P}(S_T|\tilde{y}_T)$ first. We do this by setting arbitrary values for the parameters and then using the following expression:

$$\begin{aligned} g(\tilde{S}_T|\tilde{y}_T) &= g(S_T|\tilde{y}_T) \prod_{t=1}^{T-1} g(S_t|S_{t+1}, \tilde{y}_t) \\ &= g(S_T|\tilde{y}_T) \prod_{t=1}^{T-1} g(S_{t+1}|S_t) g(S_t|\tilde{y}_t) \end{aligned}$$

Notice now that we can sample from $\tilde{S}_T$ if we have $g(S_{t+1}|S_t)$ (which is nothing more than the transition probability from one state to another) and $g(S_t|\tilde{y}_t)$ (which can be obtained from Hamilton's filter), $\forall t \in \{1, ..., T\}$.

Hamilton's filter

The basic Hamilton filter can be described as input-output-byproduct.

input: $g(S_{t-1} = s_{t-1}|\tilde{y}_{t-1})$
output: $g(S_t = s_t|\tilde{y}_t)$
byproduct: $f(y_t|\tilde{y}_{t-1})$

Running the Hamilton filter for $t \in \{1, ..., T\}$ we get the desired values $g(S_1|\tilde{y}_1), ..., g(S_T|\tilde{y}_T)$. These values are then used in the process below in order to generate $\tilde{S}_T$.

$$\mathbb{P}(S_T = i|\tilde{y}_T) = \frac{g(S_T = i|\tilde{y}_T)}{\sum_{j=1}^4 g(S_T = j|\tilde{y}_T)}$$

This probability is used to draw a sample of S_T.

$$\mathbb{P}(S_{T-1} = i|\tilde{y}_{T-1}) = \frac{g(S_{T-1} = i|\tilde{y}_{T-1})}{\sum_{j=1}^4 g(S_{T-1} = j|\tilde{y}_{T-1})}$$
$$= \frac{g(S_T|S_{T-1} = i) g(S_{T-1} = i|\tilde{y}_{T-1})}{\sum_{j=1}^4 g(S_T|S_{T-1} = j) g(S_{T-1} = j|\tilde{y}_{T-1})}$$

The above probability and the the previously simulated S_T are used to simulate S_{T-1}.

$$\vdots$$

$$\begin{aligned} \mathbb{P}(S_1 = i|\tilde{y}_1) &= \frac{g(S_1 = i|\tilde{y}_1)}{\sum_{j=1}^4 g(S_1 = j|\tilde{y}_1)} \\ &= \frac{g(S_2|S_1 = i) g(S_1 = i|\tilde{y}_1)}{\sum_{j=1}^4 g(S_2|S_1 = j) g(S_1 = j|\tilde{y}_1)} \end{aligned}$$

Using the S_2 and the above expression we can simulate S_1 which gives us the last component of $\tilde{S}_T$. This means that for every $t \in \{1, ..., T\}$ we know what the distribution of y_t is since we know which state we are in.

6 Case Study

Our case study deals with applying the above theory to an indicator that would play a role similar to that of the VIX index. The data set we will be using is the set of S&P500 weekly prices while the chosen time interval is 3 January, 2007 to 29 December, 2014. We picked this interval to include the sub-prime mortgage crash of 2008 as well as the subsequent period of relative calm in order to see how the model performs in both situations. In particular we improve on [1], see also [2] where the model was very effective in periods of high volatility but much too smooth when volatility was low. Our results are summarized below.

Parameter	Estimated Value
λ	0.00252163
γ_1^2	0.1197211
γ_2^2	0.2479246
γ_3^2	0.4250766
γ_4^2	1.370165

We estimate the variance of $y_t|\lambda$ using:

$$\begin{aligned} \mathbb{E}[\lambda\gamma_t^2|\psi_t] = \lambda \Big(&\mathbb{P}(S_t = 1|\psi_t)\gamma_1^2 + \mathbb{P}(S_t = 2|\psi_t)\gamma_2^2 \\ &+ \mathbb{P}(S_t = 3|\psi_t)\gamma_3^2 + \mathbb{P}(S_t = 4|\psi_t)\gamma_4^2 \Big) =: \hat{\sigma}_t \end{aligned}$$

We now compare $\hat{\sigma}_t$ to the VIX index.

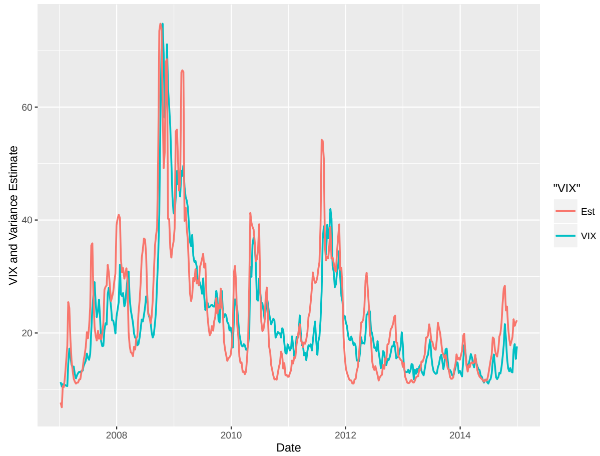

Figure 1: Visual comparison between the VIX and the expected standard deviation. We have applied a linear scaling function.

Looking at the graph we can be quite happy with the result. The estimate catches the peaks well although it tends to overshoot a bit on some of them. The problem of estimator smoothness has been solved and the way in which the estimated standard deviation mimics the VIX when there is low volatility in the market is satisfactory.

7 Conclusion and Future Developments

In this paper we applied Gibbs Sampling to the problem of estimating the parameters of a Markov Switching Model. Although most papers model the data using a Normal distribution we decided to use α-stable distributions in order to model the fat tails that financial data exhibit. We got around the problem of not having a closed form of the density of an α-stable distribution by representing it as a conditionally Normal distribution. This form was sufficient for the Gibbs Sampling approach since it utilizes conditional distributions for sampling.

Regarding future developments there are multiple ways to improve the model. The main point of interest are the transition probabilities, which in this model are in their simplest form. Using transition probabilities that that depend on multiple past states, that depend on the length of time the data has been in a certain state or that depend on other observable variables are all things that should be tried in order to make this model more accurate and robust.

References:

[1] Di Persio, L. , Frigo, M. ”Gibbs sampling approach to regime switching analysis of financial time series.” Journal of Computational and Applied Mathematics (2016).

[2] Di Persio, L. , Frigo, M. ”Maximum Likelihood Approach to Markov Switching Models.”, WSEAS Transactions on Business and Economics, , Volume 12, 2015, Art.21, pp. 239-242

[3] Salas-Gonzalez, D., Kuruoglu, E. E. , Ruiz, D. P. ”Modelling with mixture of symmetric stable distributions using Gibbs sampling.” Signal Processing (2010).

[4] Hamilton, J. D. ”A new approach to the economic analysis of nonstationary time series and the business cycle.” Econometrica: Journal of the Econometric Society (1989).

[5] Kim, C-J. , Nelson, C. R. ”State-space models with regime switching: classical and Gibbs-sampling approaches with applications.” MIT Press Books 1 (1999).

[6] Carlin, B. P. , Louis, T. A. ”Bayesian Methods for Data Analysis, Third Edition.” Chapman & Hall/CRC Texts in Statistical Science (2009).

[7] Samorodnitsky G., Taqqu M. S. ”Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.” Chapman & Hall/CRC (1994).

[8] Chauvet M. ”An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching.” International Economic Review (1998).