Maximum likelihood approach to Markov switching models

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

MATTEO FRIGO
University of Verona
Department of Computer Science
Strada Le Grazie 15, 37100 Verona
ITALY
matteo.frigo@outlook.com

Abstract: The present paper concerns a *Maximum Likelihood* analysis for the Markov switching approach to the forecasting problem of financial time series. In particular we model the volatility parameter characterizing time series of interest as a state variable of a suitable Markov chain. Latter formulation is based on the idea of describing abrupt changes in the behaviour of studied financial quantities due to, e.g., social or political factors able to substantially change the economic scenarios we are interested in. A case study for the NASDAQ IXIC index in the period 3rd Jan 2007 - 30th Dec 2013 is also provided.

Key–Words: Maximum Likelihood , Markov Switching , State Space Model

1 Introduction

Natural, physical and financial phenomena can be efficiently described as dynamical systems whose behaviours can rarely be defined by deterministic laws. Difficulties grow when we try to extrapolate information concerning a specific phenomenon solely by using empirical data. Latter approach can be successfully followed, e.g., in studying the Carbon-14 decay, which follows the differential law

$$y'(t) = -ky(t), \quad t \in [0, +\infty) = \mathbb{R}^+, \tag{1}$$

k being the *decay constant*, while $y(t)$ is the concentration of Carbon-14 and k, y are the *observable variables*. Nevertheless latter scenario is not feasible in most of the case since the dynamics in which we are interested are governed by *unobservable variables* characterizing the peculiarities of the particular studied scenario. Such a result of our ignorance about describing the dynamic of our interest is the core of many research areas like pattern recognition, financial forecasting, quantum mechanics, population dynamics, interacting particle systems, etc.

In this work we are going to deal with *discrete state-space models* (DSSM) in the financial framework in order to efficiently forecasting economical quantities starting by the study of related time series $\{y_t\}_{t=1}^T$, $1 < T \in \mathbb{N}^+$ being a certain *expiration date* or, simply, the maximum length of the series we can achieve, e.g., from the market indexes. In particular we are interested in system of the following form:

$$\begin{cases} y_t = f(S_t, \theta, \psi_{t-1}) \\ S_t = g\left(\tilde{S}_t, \psi_{t-1}\right) \\ S_t \in \Lambda \end{cases} \tag{2}$$

where $\psi_t := \{y_k : k = 1, \dots, t\}$, Λ represents the set of the all the possible states, g is the switching-state law, namely a function of both the past states and the observed data, and f can be thought as the *black-box* realizing the value of the time series at time t with respect to the information carried by S_t, ψ and θ, representing the set of the *descriptive parameters*.

2 Regime Switching

A crucial point in the DSSM theory concerns the identification of the transition law between states, namely the function g in (2). From our financial perspective, such a law has to be capable of capture and quantify structural break which have become quite common in finance, particularly due to the extreme financialization of derivatives and related markets as well as concerning rapid changing of political scenarios in most of the so-called emerging economies whose behaviour deeply influences many aspects of the worldwide economy. If the switching between two different financial states happens, e.g., at time $t+1$, for a given $t \in \mathbb{N}^+$, being determined only by the state S_t of the system at the immediate preceding time t, non matter about what is happened in the past, namely in the discrete time interval $\{0, 1, \dots, t-1\}$, we say that S_t

is a Markov process and we are in the framework of *Markov Switching*. In particular the switching law can be described by a time homogeneous Markov chain whit transition matrix $P := \{p_{ij}\}_{i,j\in\Omega}$, Ω being the set of all the possible states for the system.

In what follows we explicitly consider the case of a *first order autoregressive* pattern with regime switching, the latter being the standard choice in financial time series analysis, and we extend this approach considering a state variable for the mean and the variance of the error term, namely we study the following system:

$$\begin{cases} y_t &= \phi_1 y_{t-1} + \mu_{S_t} + \sigma_{S_t} z_t, \\ z_t &\sim \mathcal{N}(0,1), \\ S_t &\in \Omega = \{1,\dots,M\}, \\ p_{ij} &= \mathbb{P}(S_t = j | S_{t-1} = i), \\ \pi_0 &= [\mathbb{P}(S_0 = 1), \dots, \mathbb{P}(S_0 = M)], \\ \sigma^2_{S_t} &= \sigma^2_j \quad \text{if } S_t = j \quad \forall j = 1,\dots,M, \\ \mu_{S_t} &= \mu_j \quad \text{if } S_t = j \quad \forall j = 1,\dots,M. \end{cases} \tag{3}$$

Our model structure belongs to the *Gaussian State-Space models* family and it is widely used in many research topics such, e.g., in population dynamics, see [6], or for wind forecasting, see [1].

3 Filtering

The system described by eq. (3) is non-linear in the parameter S_t, so we can not exploit the usual tools of inference related to Linear Gaussian DSSM, namely the *Kalman* filter. An alternative is given by the *Hamilton* filter, a modification of the Kalman approach, proposed by Hamilton in 1989, see [3]. Following latter approach, we suppose to be able to estimate the state probabilities $\mathbb{P}(S_t = j|\psi_t)$, the transition probabilities p_{ij} and the steady-state probability $\pi_0 = [\mathbb{P}(S_0 = 1|\psi_0), \dots, \mathbb{P}(S_0 = M|\psi_0)]$, where ψ_0 is a formal expression for the state value at the initial time, while ψ_t is the observed time series until time t, namely $\psi_t = \{y_\tau : \tau = 1,\dots,t\}$. We also consider the marginal distribution of y_t, conditional to S_t, S_{t-1} and ψ_{t-1}, to be

$$f(y_t|S_t, S_{t-1}, \psi_{t-1}) = \frac{e^{\frac{\left(y_t - \mu_{S_t} - \phi\left(y_{t-1} - \mu_{S_{t-1}}\right)\right)^2}{2\sigma^2_{S_t}}}}{\sqrt{2\pi\sigma^2_{S_t}}} \tag{4}$$

Examples that show how to compute aforementioned objects can be found, e.g., in citeDiP14 and [5]. From a concrete point of view it is useful to describe the Hamilton filter exploiting the following operative *recipe*

Filtering procedure

1. Compute the transition probabilities $P = p_{ij}$.
2. Compute the *steady-state* probabilities.
3. Perform the following steps `for` $t = 2,\dots,T$:
   (a) Compute the probability of each state j at time t conditional to the dataset ψ_{t-1}:
   $$\mathbb{P}(S_t = j) = \sum_{i=1}^{M} p_{ij}\mathbb{P}(S_{t-1} = i|\psi_{t-1}). \tag{5}$$
   (b) Compute the joint density of y_t, S_t and S_{t-1} given ψ_{t-1} and its marginal density with respect to y_t:
   $$f(y_t, S_t, S_{t-1}|\psi_{t-1}) = f(y_t|S_t, S_{t-1}, \psi_t)\,\mathbb{P}(S_t = j, S_{t-1} = i|\psi_{t-1}), \tag{6}$$
   and
   $$f(y_t|\psi_{t-1}) = \sum_{j=1}^{M}\sum_{i=1}^{M} f(y_t|S_t, S_{t-1}, \psi_{t-1})\,\mathbb{P}(S_t = j, S_{t-1} = i|\psi_{t-1}). \tag{7}$$
   (c) Update the joint probabilities of S_t and S_{t-1} given ψ_t:
   $$\mathbb{P}(S_t = j, S_{t-1} = i|\psi_t) = \frac{f(y_t, S_t = j, S_{t-1} = i|\psi_t)}{f(y_t|\psi_{t-1})} \tag{8}$$
   (d) Compute the updated probabilities of S_t given ψ_t for all $j = 1,\dots,M$:
   $$\mathbb{P}(S_t = j|\psi_t) = \sum_{i=1}^{M}\mathbb{P}(S_t = j, S_{t-1} = i|\psi_t). \tag{9}$$

3.1 Smoothed probabilities

The output of the previously determined *filtering procedure* consists in the probabilities set of being in a state S_t conditioned on the past or at most on the past history of the process plus its present values. Sometimes it could be also interesting to compute the probabilities conditioned on the whole time series, e.g., if we wanted to write an expression for

$$\mathbb{E}\left(\mu_t|\psi_T\right) = \sum_{j=1}^{M} \mu_j \mathbb{P}\left(S_t = j|\psi_T\right) \,, \tag{10}$$

$$\mathbb{E}\left(\sigma_t^2|\psi_T\right) = \sum_{j=1}^{M} \sigma_j^2 \mathbb{P}\left(S_t = j|\psi_T\right) \,, \tag{11}$$

and, in order to do that, we need to exploit a *smoothing algorithm*, which is a backward procedure that returns the set $\{\mathbb{P}\left(S_t = j|\psi_T\right) : t = 1, \dots, T\}$, see, e.g., [2, 4, 5], for details, while in what follows we will only give a sketch of the related procedure.

1. Consider the fact that in the framework of DMSM the knowledge of y_{t+1} does not provide more information about S_t than $\{S_{t+1}, \psi_t\}$, at least if the data are uncorrelated or at most first-order autoregressive.

2. Prove the following formula, see, e.g., [2]:

$$\mathbb{P}\left(S_t = i|S_{t+1} = j, \psi_T\right) = \mathbb{P}\left(S_t = i|S_{t+1} = j, \psi_t\right). \tag{12}$$

3. Exploiting the Bayes' rule, show that the following relation holds:

$$\mathbb{P}\left(S_t = i, S_{t+1} = j|\psi_T\right) = \mathbb{P}\left(S_{t+1} = j|\psi_T\right) \frac{\mathbb{P}\left(S_t = i, S_{t+1} = j|\psi_t\right)}{\mathbb{P}\left(S_{t+1} = j|\psi_t\right)}. \tag{13}$$

4. Compute the needed quantity :

$$\mathbb{P}\left(S_t = i|\psi_T\right) = \sum_{j=1}^{M} \mathbb{P}\left(S_t = i, S_{t+1} = j|\psi_T\right). \tag{14}$$

4 Identification

The next problem we have to face consists in identifying the parameters μ, σ and p_{ij} of the system (3). Using the standard approach provided by maximum likelihood technique, we can exploit the *Maximum-Likelihood Estimation* (MLE) which allows us to rephrase the problem in terms of the optimization of the functional

$$\ln \mathcal{L} := \sum_{t=1}^{T} \ln\left(f_{y_t|\psi_{t-1}, S_t, S_{t-1}}\right) \,. \tag{15}$$

Since we are dealing with a recursive filter, we just need to insert the identification step at some point in the filtering procedure described above. In particular, we start with the initialization of the log-likelihood function $\ell\,(t = 0) = 0$ and after the 3b-step, we update it computing

$$\ell\,(\theta, \psi_t) = \ell\,(\theta, \psi_{t-1}) + \ln\left(f\left(y_t|\psi_{t-1}\right)\right), \tag{16}$$

then we maximize it with respect to

$$\theta_t = [\mu_1, \dots, \mu_M, \sigma_1, \dots, \sigma_M]_t \tag{17}$$

under the constraint $\sigma_1 < \cdots < \sigma_M$, and we use the estimated θ for the last steps computations of the filtering procedure.

5 Case study

In what follows we shall apply previously obtained results considering the weekly returns of the NASDAQ index in the period 3rd Jan 2007 - 30th Dec 2013. Note that latter time series has zero-mean, hence the model is reduced to a switching variance problem. Moreover the related AR(1) coefficient is of order $\phi \propto 10^{-2}$, therefore we can take into account the serially uncorrelated version of the Hamilton filter. We have chosen to implement a three states DMSM since this is the standard choice in financial applications, see, e.g., [5] and the obtained results are given by

Parameter	Est.Value
σ_1^2	$1.4865e-04$
σ_2^2	$4.1009e-04$
σ_2^3	$9.0873e-04$

with transition matrix

$$P = \begin{bmatrix} 0.96 & 0.19 & 0.00 \\ 0.00 & 0.81 & 0.09 \\ 0.04 & 0.00 & 0.91 \end{bmatrix}. \tag{18}$$

We want to test the goodness of our fit through the analysis of the standardized residuals. From the assumptions of our model we would like to observe that they follow a Gaussian distribution, hence we start with a normal probability plot for a visual analysis, see figure 1. We see that there are some outliers that could lead to the failure of a normality test, but the elimination of the biggest two of them lets us to not reject the normality test with a significance level of the 5%. We also report the plot of the smoothed probabilities of the three states in figure 2 and the conditional standard error in figure 3. The first graph shows how the high volatility state has a cyclic nature, which corresponds to the fact that great changes within the information technology economy are usually shown in the same period of the year. The second plot represents the implied volatility, which is a measure of how much we expect the price of the index differs from the previous value, namely it constitutes a common risk representation widely used by traders.

6 Conclusion

The plot of the implied volatility and of the smoothed probabilities are compatible in the highly risky states identifications. Moreover we had only a 0.5% of non significant data, being obliged to cut only two values over 365 to satisfy the normality test. It follows that our model provides a concrete and affective description of the risk related to investments concerning the NASDAQ IXIC index.

Acknowledgements: We would like to thank Marcelo Perlin for the *MS Regress* package, see [7], and Dr. Samuele Vettori for fruitful suggestions and comments.

References:

[1] Ailliot, Pierre et al. ”Non-homogeneous hidden Markov-switching models for wind time series.” Journal of Statistical Planning and Inference (2014).

[2] Di Persio, Luca and Vettori, Samuele. ”Markov Switching Model Analysis of Implied Volatility for Market Indexes with Applications to S&P 500 and DAX.” Journal of Mathematics 2014 (2014).

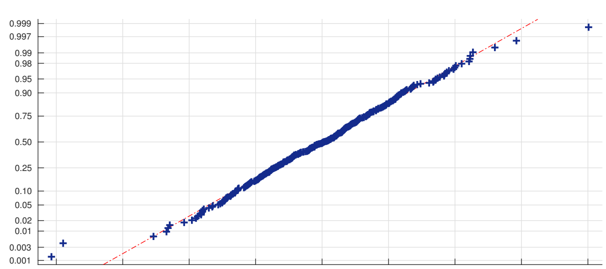

Figure 1: Normal Probability Plot of the standardized residuals.

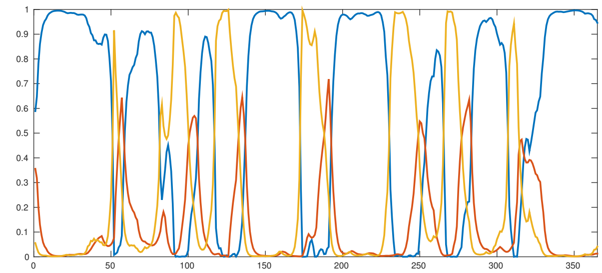

Figure 2: Blue = state 1, red = state 2, yellow = state 3.

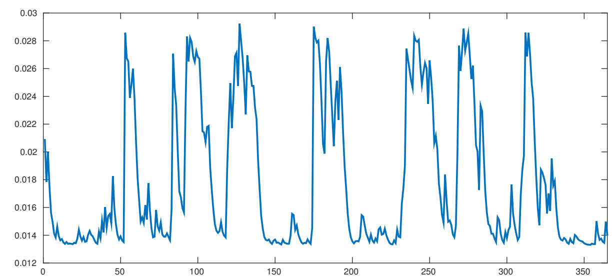

Figure 3: Conditional standard error.

[3] Hamilton, James D. ”A new approach to the economic analysis of nonstationary time series and the business cycle.” Econometrica: Journal of the Econometric Society (1989): 357-384.

[4] Hamilton, James D. Time series analysis. Vol. 2. Princeton: Princeton university press, 1994.

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

[6] Luo, Qi and Xuerong, Mao. ”Stochastic population dynamics under regime switching.” Journal of Mathematical Analysis and applications 334.1 (2007): 69-84.

[7] Perlin, Marcelo. ”MS_Regress-the MATLAB package for Markov regime switching models.” Available at SSRN 1714016 (2012).