Liu and West filter for parameter estimation of Markov Swithing Multifractal Model

Jan Greve[a], Jaeyeon Lee[b], Kangcheng Lin[c]

[a] *Duke University Department of Statistical Science*
[b] *Duke University Department of Statistical Science*
[c] *Duke University Department of Statistical Science*

Abstract

Markov Switching Multifractal Model is a stochastic volatility model used in finance based on an assumption that the return volatility is a multifractal system. This assumption was first introduced by Mandelbrot who coined the word fractals and considered finance as one area of application for this concept. This lead him to create the Multifractal Model of Asset Returns (MMAR)[1] but did not gain wide recognition due to its combinatorial nature being impractical in many application settings. Markov Switching Multifractal (MSM) by Calvet et al [2][3] is a reinterpretation of MMAR in state space model (or hidden Markov model) which already has a significant foothold in areas such as econometrics, finance and statistics. Due to this nature, MSM has gained popularity among finance researchers and is widely applied to model volatile assets such as commodity and fx futures.

Depending on the specification, MSM can either have finite or infinite number of states. Calvet et al [3] considered the finite case and employed maximum likelihood estimation to estimate parameters of this model by taking into account all possible paths from first order Markov chain and picked the set of parameters with the highest likelihood. This approach is computaionally expensive and will be infeasible quickly when the number of states are increased to enhance its predictive capability.

In this report, we instead considered applying particle filter to approximate the likelihood and get an estimate of optimal set of parameters. Specifically, we employed Liu and West filter[4] and examined its performance in relation to the MLE approach with simulated MSM process.

Keywords: Markov Swithing Multifractal, particle filter, Liu and West filter

Email addresses: jan.tomoya.greve@duke.edu (Jan Greve), jaeyeon.lee@duke.edu (Jaeyeon Lee), kangcheng.lin@duke.edu (Kangcheng Lin)

1. Model Description

MSM is a stochastic volatility model developed by Calvet and Fisher[3] and is often denoted as MSM($\bar{k}$) because it consists of $\bar{k}$ multipliers

$$M_t = (M_{1,t}, ..., M_{\bar{k},t}) \in \mathbb{R}_+^{\bar{k}}.$$

Each components have the same marginal distribution $M(\theta)$, but the switching frequency is given by exogenous parameters $\gamma \equiv (\gamma_1, \gamma_2, ..., \gamma_{\bar{k}})$. Therefore the dynamics of multipliers are specified as following

$$M_{k,t} = \begin{cases} m \sim M(\theta) & \text{with probability } \gamma_k \\ M_{k,t-1} & \text{with probability } 1 - \gamma_k \end{cases}$$

where the switching probability γ_k is specified as

$$\gamma_k = 1 - (1 - \gamma_1)^{b^{k-1}}.$$

Due to this construction, multipliers are mutually independent with each other and the stochastic volatility can be considered as a function of the vector M_t with first-order Markov property.

$$\sigma(M_t) \equiv \bar{\sigma}\Big(\prod_{i=1}^{\bar{k}} M_{k,t}\Big)^{\frac{1}{2}},$$

where $\bar{\sigma}$ is a positive constant to be estimated. Finally, returns r_t are

$$r_t = \sigma(M_t)\epsilon_t \quad \text{where } \epsilon_t \sim N(0,1).$$

which is just a realization of the state $\sigma(M_t)$ with the Gaussian error.Thus,the full parameter vector is $\psi \equiv (\theta, \bar{\sigma}, b, \gamma_1)$. For the marginal distribution $M(\theta)$, any distribution with positive support will do the job as long as $\mathbb{E}(m) = 1$.

Essentially, MSM($\bar{k}$) is a HMM with $\bar{k}$ components and if we use the simplest specification of $M(\theta)$ where m only takes two values with equal probability, the transition matrix of the state M_t is $2^{\bar{k}} * 2^{\bar{k}}$. The figure in the next page is a graphical representation of this HMM (grey nodes are non-stochastic).

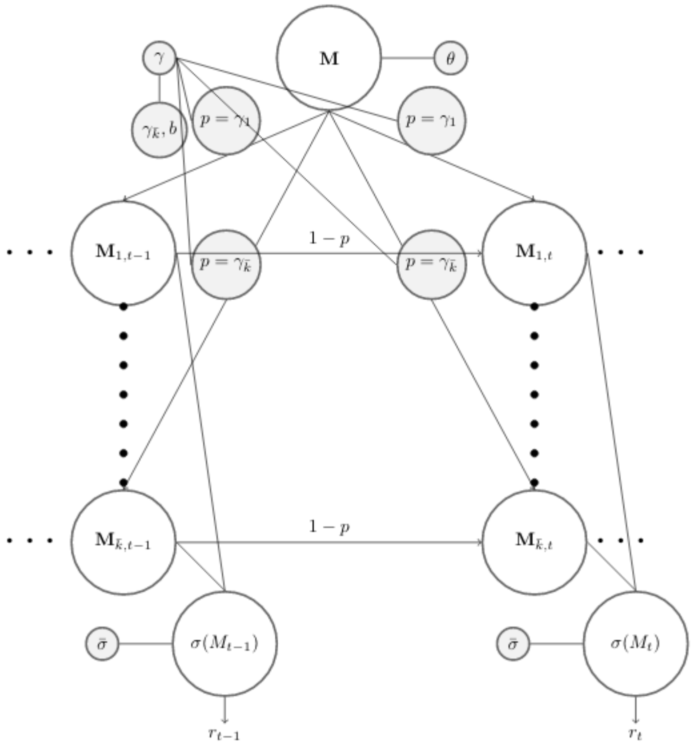


Figure 1: Graphical representation of MSM($\bar{k}$)

2. Particle filtering for simulating likelihood

2.1. Proof of our particle filter implementation to $MSM(\bar{k})$

The following proof provides the justification of the particle filtering we used in the code for approximating the posterior distribution of the states $\Pi_{t+1} \in \mathbb{R}_+^{2^{\bar{k}}}$ with two stage M_{T+1} sampling. Let F a function that maps $\mathbb{R}$ to $\mathbb{R}_+^{\bar{k}}$.

Let d be the number of possible values that m^j can take and $R_{t+1} \equiv \{r_s\}_{s=1}^{t+1}$. Let $h(m) \equiv P(M_{t+1}|R_t)$

$$
\begin{aligned}
E[F(M_{t+1})|R_{t+1}] &= \sum_{j=1}^{d} P(M_{t+1} = m^j|R_{t+1})F(m^j) \\
&= \sum_{i=1}^{d} h(m^j)\frac{P(M_{t+1} = m^j|R_{t+1})}{h(m^j)}F(m^j) \\
&\because h(m^j) = P(M_{t+1} = M^j|R_t), \\
&\text{by the convergence of the Monte Carlo Integration,} \\
E[F(M_{t+1})|R_{t+1}] &\approx \frac{1}{B}\sum_{b=1}^{B} \frac{P(M_{t+1} = \hat{M}_{t+1}^{(b)}|R_{t+1})}{h(\hat{M}_{t+1}^{(b)})}F(\hat{M}_{t+1}^{(b)}) \\
&\text{by Bayes Rule, the approximation above turns out to be importance sampling} \\
&\because \frac{P(M_{t+1} = \hat{M}_{t+1}^{(b)}|R_{t+1})}{Bh(\hat{M}_{t+1}^{(b)})} = \frac{f_{r_{t+1}}(r_{t+1}|M_{t+1} = \hat{M}_{t+1}^{(b)})}{Bf_{r_{t+1}}(r_{t+1}|R_t)} \\
&\text{Moreover, the denominator can be approximated as follows} \\
f_{r_{t+1}}(r_{t+1}|R_t) &\approx \frac{1}{B}\sum_{b=1}^{B} f_{r_{t+1}}(r_{t+1}|\hat{M}_{t+1}^{(b)}) \\
&\therefore E[F(M_{t+1})|R_{t+1}] \approx \sum_{b=1}^{B} \frac{f_{r_{t+1}}(r_{t+1}|M_{t+1} = \hat{M}_{t+1}^{(b)})}{\sum_{b=1}^{B} f_{r_{t+1}}(r_{t+1}|\hat{M}_{t+1}^{(b)})}F(\hat{M}_{t+1}^{(b)}) \\
&= \sum_{b=1}^{B} W_b F(\hat{M}_{t+1}^{(b)})
\end{aligned}
$$

As shown above, The expectation $E[F(M_{t+1})|R_{t+1}]$ can be approximated with the weighted average of the statistics of drawn samples, weighted by the likelihoods of the returns. This resembles the form of approximation using importance sampling. Since, F function was defined as a mapping from the real number space to the positive real value space, F can be replaced with the state conditional probability Π_{t+1}. Therefore, the proof demonstrates that the code implementation of our particle filtering can be justfied to approximate Π_{t+1} with re-sampled $\{\hat{M}_{t+1}^{(i)}\}_{i=1}^{B}$

2.2. Procedure of Particle Filtering

At fixed time t, i) Draw $\{M_{t+1}^{(i)}\}_{i=1}^{B}$ using $P(M_{t+1}^{i}|M_t^i)$(Transition Matrix). ii) Reweight the probability of draws using importance sampling and re-draw $\{\hat{M}_{t+1}^{(i)}\}_{i=1}^{B}$ using those reweighted probabilities.

Assume $\{M_t^{(i)}\}_{i=1}^{B}$ have been independently drawn from Π_t

Π_{t+1} can be updated given a new return r_{t+1} as follows. $\Pi_{t+1}^{j} \propto f_{r_{t+1}}(r_{t+1}|M_{t+1} = m^j)\sum_{i=1}^{d} P(M_{t+1} = m^j|M_t = m^i)\Pi_t^i$

In the first step, draw $M_{t+1}^{(i)}$ given $M_t^{(i)}$ using one-step ahead Marcov chain property of M_t Repeat this B times, then first stage of $\{M_{t+1}^{(i)}\}_{i=1}^{B}$ are obtained. However, the samples drawn given M_t uses information available up to time t only. Therefore, it is necessary to update M_{t+1} incorporating the information of the return at $t+1$ in the following step.

3. Liu and West filter

Particle filter is a useful tool in approximating the state evolution at any given time and obtain an approximated likelihood of a model with given parameters at the same time. However, there are several issues in obtaining the MLE of simulated likelihood obtained from generic particle filters which we will detail below. For this reason, we won't be able to pick the optimal set of parameters by simply maximizing the simulated likelihood.

Therefore, we employ Liu and West filter (from here on we call it L-W filter) which tackles this issue.

3.1. Problems of using generic particle filter in parameter estimation

The likelihood of a parameter θ given observations y and the predictive distribution $p(y_t|y_{1:(t-1)}, \theta)$ at time t is obtained as follows

$$L(\theta|y_{1:t}) = \prod_{t=1}^{T} p(y_t|y_{1:(t-1)}, \theta).$$

A particle filter approximation for this likelihood is

$$L(\theta|y_{1:t}) \approx \prod_{t=1}^{T} \frac{1}{M} \sum_{i=1}^{M} \omega_t^i.$$

Therefore, the log likelihood is given as

$$L(\theta|y_{1:t}) \approx \sum_{t=1}^{T} \log\left(\sum_{i=1}^{M} \omega_t^i\right) - T\log(M).$$

This approximation contains a Monte Carlo error which results in highly wiggly surface to obtain a minimum unless we increase the sample size M which will lead to another problem of an increase of computational complexity.

Moreover, since the value consists of weights ω_t^i instead of θ, we won't be able to obtain a gradient, thus Newton method and its variations are not applicable for this task.

For these reasons, it is highly unlikely to obtain a reliable MLE from simulated likelihood of a generic particle filter.

3.2. Self organizing state space model

To circumvent this issue, Kitagawa[5] suggested a new specification called self organizing state space model.

In Bayesian statistics, both the state x_t and the parameters θ has its own distribution. Thus, we can instead consider the joint distribution $p(x_t, \theta|y_{1:t})$ as a state. Then, after obtaining the posterior joint distribution with particle filtering, we can marginalize one from the other to obtain the posterior for each measurements.

However, for this approach to be successful, the particles of θ at time 0 (so the particles generated from the prior of θ) must contain the true parameter θ in its range. An appropriate specification of such prior is not trivial (especially when the dimension of θ is big).

Instead, Kitagawa[5] treated θ to also be time dependent along with x_t with some innovations so that the particle filter simulation will guide θ_t towards its more likely position.

3.3. L-W filter

Liu and West[4] expanded this idea by introducing kernel smoothing which sets the distribution of θ_t conditioned on the observation at time $t-1$ as follows

$$p(\theta_t|y_{t-1}) \sim N(a\theta_{t-1} + (1-a)\bar{\theta}_{t-1}, (1-a^2)V_{t-1})$$

where $a = \frac{(3\delta-1)}{2\delta}$ with δ suggested to take values between 0.95 to 0.99.

This filter when t is sufficiently large will have V_t converging to 0. However, sometimes when θ_t is too distant from the true θ, it will fail to achieve convergence (which we will show in a later chapter). To avoid this, we must carefully choose the prior distribution of θ_t in order to prevent overwhelming number of samples to be generated in an area with low likelihood.

4. Implementation of L-W filter

As with regular particle filter, initialization of a L-W filter is a nontrivial task. In this chapter, we discuss several steps we have taken so that L-W filter will converge to the target parameters and show an example where it didn't to highlight some of the pitfalls it has.

4.1. Initialization

There are several potential issues regarding the initialization of parameters and distributions. We will discuss some of those we have encountered through experimenting with the simulated data.

4.1.1. Bad initialization of parameter values

L-W filter parameter estimate will fail to converge to the true parameter (by construction it will still converge to somewhere) values if initialization of parameters are sufficiently far away from the true values. This is because the conditional evolution of the parameter follows a normal distribution which does not have fat tails, and its variance is decreasing by time.

(a) a = 0.99

(b) a = 0.9

Figure 2: Examples of L-W filter failing to converge to the true b = 5 with different value of a (other values are identical)

In our experiment, we tried to circumvent this issue by increasing the particle size so that some particles will be generated from a place close the true parameter value and the resampling will draw the mean of the distribution towards that direction. However, in most cases even after the resampling, the weighted mean was overwhelmed by particles from distant values.

Another adjustment we tried was to adjust the value of a which controls the spread and the overall closeness of particles to its weighted mean. When the parameter a is set close to 1 (like 0.99) it will allow individual particles to have values diverting from overall weighted mean so that it can explore more distant areas but at the cost of increased variance and slow convergence.

In our case, this lead the filter to converge to a distant point which even with resampling did not revert back to more likely areas since by construction, weighted variance will decrease by time. The figure above is a result of one parameter which experienced this issue.

By lowering the value of a, we can prevent the sharp jump in Fig2 to happen since particles will be more concentrated to its weighted mean. However, then when a drift happens it will be less likely to be corrected since most particles are close to each other so weights are more or less uniform even though it is totally missing the true parameter.

4.1.2. Bad initialization of sampling distributions

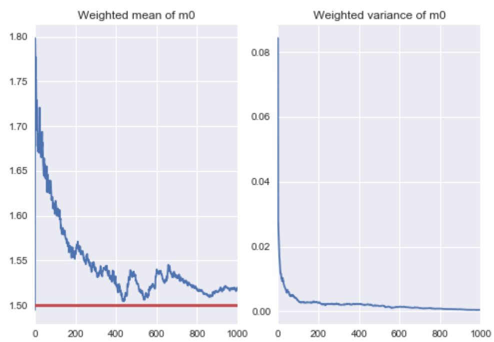
The aforementioned issue lead us to believe that the time evolution of parameters may not necessarily be given a normal distribution. Instead, we might consider using Cauchy distribution which has fatter tails than the normal.

4.2. Successful case of L-W filter and its potential compared to the MLE approach

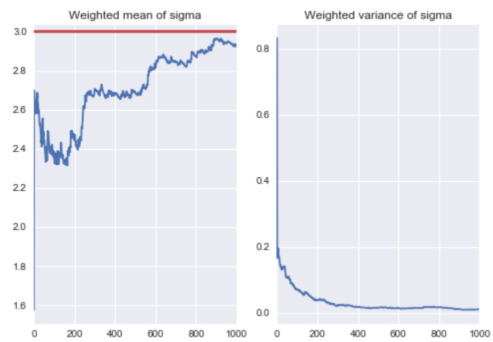
When the initialization is appropriately done, L-W filter will converge to the neighbourhood of the true value of the parameter. Here, we show the successful implementation of L-W filter to the simulated $MSM(3)$ data with $b = 0.5, \theta = m_0 = 1.5, \bar{\sigma} = 3, \gamma_{\bar{k}} = 0.5$.

4.2.1. Simulation results

We run L-W filter (the code we wrote is given in the appendix) with $a = 0.95$, 10000 particles.

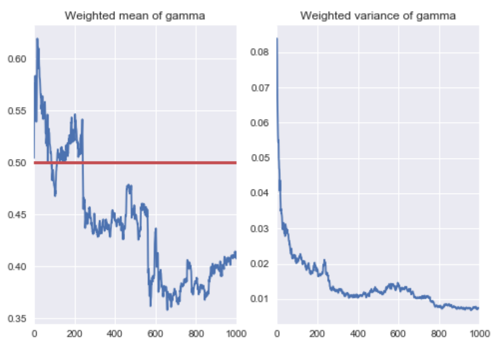


(a) Convergence to $m_0 = 1.5$



(b) Convergence to $\bar{\sigma} = 3$

(c) Convergence to $b = 5$



(d) Convergence to $\gamma_{\bar{k}} = 0.5$

Figure 3: Converngence of L-W filter to true parameter values

4.2.2. Advantages of L-W filter compared to the MLE approach

Although we highlighted several issues regarding the L-W filter, it is fundamentally more efficient and superior to MLE approach in several respect.

To begin with, once a new observations is obtained, the MLE approach must recompute the likelihood using all past data while L-W filter is updated sequentially, thus will require minimum amount of computation. It is true that taking several thousands of samples is also computationally expensive but that is parallelizable.

Secondly, the MLE approach requires the full transition matrix of M_t which in binomial MSM is $2^{\bar{k}} * 2^{\bar{k}}$ matrix. For this reason, when $\bar{k}$ is set to a larger value it is computationally infeasible to calculate and store the whole matrix since the complexity is squared of $2^{\bar{k}}$. The L-W filter on the other hand will only require one row of the matrix for each particles, so the computational complexity is linear to $2^{\bar{k}}$.

Finally, for this report we dealt with binomial MSM where the distribution M takes only two values. In fact the MSM sets almost no restriction to the distribution of M other than its mean to be 1 and that it will only take positive values. Threfore when we consider the case when M is truncated normal for example, the MLE approach will fail since we don't have a finite transition matrix.

Appendix A. All codes we made

All codes we made for this project is available at a github repository below.
`https://github.com/Jantg/MSM_python`

References

[1] B. Mandelbrot, A. Fisher, L. Calvet, A multifractal model of asset returns., Cowles Foundation Discussion Paper (1164).

[2] L. Calvet, A.Fisher, Forecasting multifractal volatility., Journal of econometrics 105 (1) (2001) 27–58.

[3] L. Calvet, A.Fisher, How to forecast long-run volatility: regime switching and the estimation of multifractal processes., Journal of Financial Econometrics 2 (1) (2004) 49–83.

[4] L. Jane, M. West, Combined parameter and state estimation in simulation-based filtering, in: Sequential Monte Carlo methods in practice, Springer, 2001, pp. 197–223.

[5] G. Kitagawa, A self-organizing state-space model, Journal of the American Statistical Association (1998) 1203–1215.