Stock Index Estimation of HMM-SV Model

Xinyi Wu,[1*] Prof. Ying Wu[2]

[1 2]Department of Electrical Engineering and Computer Science, Northwestern University,
630 Clark Street, Evanston, IL 60208, USA
[*]To whom correspondence should be addressed; E-mail: xinyiwu2019@u.northwestern.edu.

Abstract: This work explores more possibilities to apply statistical pattern recognition techniques to financial market estimation and prediction. Based on past observations, the economists has built mathematical models to characterize some times series data, which includes the movement of stock prices, options prices, or other types of derivatives (7). In financial market, we put emphasis upon the research of volatility variations. Thus the construction of stochastic volatility models become very important. Features of stochastic volatility(SV) models are usually characterized by their different structures as well as the parameters within the same model. An heuristic method to extract features from SV models is the Hidden Markov Model(HMM). We acquire historical data as observations of the apparent chain and the model structure/parameters as hidden chain. This work will first show a general view of necessary models and finally provide details about the process of applying HMM-SV model on financial time series estimation.

1 Introduction

The idea of analyzing capital market behaviors has formed in a very early stage, and those ideas have been developing and promoting with very new technologies and statistical methods. One of the earliest mathematical models is the autoregressive conditional statistical heteroskedasticity model by Engle(1982), and later extended to Generalized (GARCH) model by Bollerslev et.al in 1995 (3). Those models successfully captures some features of financial time series. For example, Volatility Clustering: Large changes tend to be followed by large changesof either signand small changes tend to be followed by small changes (1); Continuity and Stability:continuous and stable fluctuation rate with seldom jumps, and Leverage phenomenon: volatility is different in response to a sharp rise in prices and a sharp drop in prices.

An implicated idea is then promoted: Stochastic models can be used to model the variance as an unobserved component that follows a particular stochastic process. (3) Intuitively, we know that the observation is formed by addition of some implicit process plus some random noise (This is somewhat similar to Auto-Regressive and Moving Average Model (ARMA) process). Thus a formula is defined as follow:

$$dY_t = \mu(\sigma_t^2)dt + \sigma_t dB_t$$

where $V_t = (\sigma_t)$ is the unobserved instantaneous volatility, (B_t) a Brownian motion and $\mu(\sigma_t^2)$ is some real function. (6)

Based on this model, Leroux (1992) or Bickel and Ritov (1996) come up with the Hidden Markov model, and improved by Jensen and Petersen 1998 by referring that ”the state space of hidden chain is finite” (6). To recall the ”state space”, I give a continuous time variant model as follow:

$$\begin{cases} \dot{x}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ y(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{cases}$$

$y(t)$ is refereed as the observation results and $x(t)$ are those finite states. I will expand this idea into details in the following sections.

A complete and reasonable definition of the **generalized Hidden Markov Model** is first proposed by Genon-Catalot, Jeantheau, and Laredo(2000) (6) as:

DEFINITION: A process $\{Y_n, n \geqslant 0\}$ *with state* $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ *follows a generalized hidden Markov Model with a hidden chain* $\{X_n, n \geqslant 0\}$ *if:*

(i) $\{X_n, n \geq 0\}$ *is a unobserved strictly stationary Markov chain with state space* $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

(ii) For all n, given $(X_0, X_1, \ldots, X_n)$*, the* Y_t*,* $t = 0, 1, \ldots, n$ *are conditionally independent and the conditional distribution of* Y_t *only depends on* X_t.

(iii) The conditional distribution of Y_n *given* X_n *does not depends on n*

Some notes needs to make herein. First, as I will cover in the stochastic volatility models, that the implicit process should be stationary i.e. unchanged in first and second moment, for the efficiency to calculate and modeling. ii) and iii) is actually characterizing the relationship between the observation samples and model parameters of the Hidden Markov Chains.

Started from this simple and clear, we can thereby move forward to detailed explanation. Finally, I would like to provide some state-of-the-art results that improve the accuracy of the prediction.

2 Stochastic Volatility Model

”In statistics, **stochastic volatility models** are those in which the variance of a stochastic process is itself randomly distributed.” (8) (G)ARCH model are thus regarded as the basis of the stochastic volatility model. For a long time, the econometricians and econophysicists has stated that the characterization of these models is what the price of the derivative should be as stated in book (11).

2.1 Linear Model

Assume that the time series is stationary and auto-correlated, we thus have (9)

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t + \sum_{i=1}^{q} \theta_i a_{t-i}$$

in which $\{a_t\}$ is white noise series and p, q are non-negative integer. We called it **ARMA(p,q)** Model (9).

A digression is that one of the most popular criterion to do this is the **Akaike information criterion**.

$$AIC = -2\ln(\hat{L}) + 2k$$

where k is number of parameters, $\hat{L}$ maximum value of the likelihood function for the model. AIC measures the quality of data fitting and avoid the occurrence of over-fitting in a greater probability. The best model that we should consider is the one with minimum AIC value, and researches always limit the maximum order of AR $p \leq 6$ and that of MA $q \leq 4$.

2.2 Non-linear Model

"While conventional time series and econometric models operate under an assumption of constant variance, the ***ARCH (Autoregressive Conditional Heteroskedastic)*** *process allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant."* (10) The setup of this model is :

$$\alpha_t = \sigma_t \epsilon_t$$

$$\sigma_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2$$

$$\alpha_0 > 0; \forall i > 0, \alpha_i \geq 0$$

in which $\{\epsilon_t\}$ is i.i.d random process with $E[\epsilon] = 0$ and $Var[\epsilon] = 1$, σ_t^2 is called conditional heteroskedasticity. In this model, large "perturbation" tend to follow by another large "perturbation". This is similar to the phenomenon of "volatility clustering".

When put this model into reality, researchers found that they have to reduce the lag of the volatility clustering and avoid negative variance parameters. Hence a much generalized model came out:

$$\alpha_t = \sigma_t \epsilon_t$$
$$\sigma_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$
$$\alpha_0 > 0; \forall i > 0, \alpha_i \geq 0; \beta_i \geq 0, (\alpha_i + \beta_i) < 1$$

If we replace the variable of time value to variance, it appears that the GARCH model has exactly the same structure as the ARMA model. Therefore GARCH model is actually considering the historical data of variance as well as the relationship between the variance of a random process with regard to time.

2.3 Generalized Stochastic Volatility Model

In theoretical literature, a more generalized form from GARCH for discrete-time data is proposed by Hull and White (1987),

$$\begin{aligned} X_{k+1} &= \phi X_k \sigma U_k & \qquad U_k \sim N(0,1) \\ Y_k &= \beta e^{X_k/2} V_k & \qquad V_k \sim N(0,1) \end{aligned}$$

Where $\{Y_k\}_{k\geq 0}$ are log-returns, $\{X_k\}_{k\geq 0}$ the log-volatility.$\{U_k\}_{k\geq 0}$ and $\{V_k\}_{k\geq 0}$ are independent i.i.d. sequences. β is constant scaling factor, ϕ the persistence(memory) in the volatility (how long it clusters), and σ the volatility of log-volatility, This model can also represent wide range of characteristics in financial time series. σ shows the behavior of the level of the volatility after mixing independent variables of distinct smoothness.

Moreover, by taking the square of the log returns, one can find that the stochastic volatility models actually related to conditionally Gaussian linear state-space models in such a manner:

$$\begin{aligned} X_k =& \phi X_{k-1} + \sigma U_{k-1}, \\ \log Y_k^2 =& \log \beta^2 + X_k + Z_k, \qquad\qquad where Z_k = \log V_k^2 \end{aligned}$$

Here V_k is Normal Gaussian and Z_k is log student-t distribution. With mixture of Gaussian distributions, the SV becomes a Gaussian linear state-space model, where

$$\begin{aligned} W_{k+1} =& \phi W_k + U_k, && U_k \sim N(0,1) \\ Y_k =& W_k + (\mu(C_k) + \sigma_V(C_k) V_k) && V_k \sim N(0,1) \end{aligned}$$

This representation of the stochastic volatility model may prove useful when deriving numerical algorithms to filter the hidden state or estimate the model parameters (1).

3 Mathematical Principles for Model Construction

Between Stochastic Volatility models and HMM, we have to derive an interface that connect them. In this way HMM knows the exact representation of it's basic components $\Lambda = (A, B, \pi)$

- Hidden States: $\mathcal{S} = \{S_1, \ldots, S_N | S_i \in x_t\}$

  Observation Symbols: $\mathcal{O} = \{O_1, \ldots, O_M | O_i \in z_t\}$

- Transition probability: $A = \{a_{ij}\}$

$$a_{ij} = p(x_t = S_j | x_{t-1} = S_i), \;\; 1 \leq i, j \leq N$$

- Conditional prob. of observation: $\mathbf{B} = \{b_j(k)\}$

$$b_j(k) = p(z_t = O_k | x_t = S_j), \quad 1 \leq j \leq N, 1 \leq k \leq M$$

- Prior states $\pi = \{\pi_i\}$

$$\pi_i = p(x_1 = S_i), 1 \leq i \leq N$$

.

To perform accurate estimations of the components above, I herein provide two important mathematical theories: ergodicity and ρ-mixing / β-mixing. These mathematical principles aims to provide a foundation in some revision of HMM-SV models. For a general idea of how HMM-SV model works, please refer to the next section.

3.1 Ergodic Theory

In the generalized model, $dY_t = \mu(\sigma_t^2)dt + V_t dB_t$, the unobserved quality $\{V_t\}$ is ergodic and control the volatility of (Y_t). So we are interested in studying the ergodicity of observations $(Y_{i\Delta})$(the sampled observations in a frequency $1/\Delta$). This will help us determine the Transition probability A and Conditional probability of observation b in the training of HMM model.

The ergodic theory originated from a research of the behavior of a long duration dynamical system. *A stochastic process is said to be ergodic if its statistical properties can be deduced from a single, sufficiently long, random sample of the process.* (wiki) To understand this, I assume the statistical properties is mean of a wide-sense-stationary process: $\mu_X = E(X(t))$, if the average estimate $\hat{\mu_X} = \frac{1}{T} \int_0^T X(t)dt$ converge in squared mean to the assemble average μ_X as $T\infty$, it's called ergodic. Because the estimation of the process will keep approaching the μ_X without any explosion/mutation/divergence at certain time.

Another explanation of ergodic process is that a particle will cover every state in an ensemble given enough long duration. The abstraction of ergodicity process is shown below (Just give a intuitive impression on convergence): Define domain

$$\mathcal{D} = \left\{ f \in L_\pi^2 : \left\| \frac{P_t f - f}{t} - g \right\| \to 0,\ as\, t \downarrow 0\, for\ some\ g \in_\pi^2 \right\}$$

, the infinitesimal generator and $\mathcal{A}$ is defined by

$$f \in \mathcal{D} \to \mathcal{A}\ \ f = \lim_{t \downarrow 0} \frac{P_t f - f}{t} in\ L^2_\pi$$

The remark is that the ergodicity of (X_t) is linked with the dimension of the eigenspace of $\mathcal{A}$ associated with 0. More clearly, it is that 0 is a simple eigenvalue of $\mathcal{A}$, the null space $N_{\mathcal{A}} = \{f \in \mathcal{D} : \mathcal{A}f = 0\}$ is one dimensional subspace of L^2_π spanned by constants. (2).

Mathematically we linked the ergodicity properties with calculating space $\mathcal{A}$ and checking the null space of it, which is algorithmic computable.

If the process is a time-homogeneous Markov process $(X_n, n \in \mathbb{N})$ (means that the transition probability doesn't varies with regard to time). We denote the transition probability as $P(x, dy)$ as

$$Pf(x) = \int_S f(y)P(x, dy) = E[f(X_1)/X_0 = x], \quad \pi a.s.$$

Thus the ergodicity is linked with dimension of the eigenspace of P associated with the eigenvalue 1. (2)

3.2 Mixing Theory

More statistical information on ergodicity analysis needs to depend on β-mixing or rho-mixing. Before investigating more mathematical principals about mixing, here I provide my notes on studying mixing theories on wikipedia.

We can think of mixing as a process to increase entropy in an ensemble, like mixing pure coco and pure milk to perform a "complicated" beverage - chocolate milk. In general, I regard mixing as an irreversible thermodynamics.

We can extend the idea of mixing into stochastic process theory. A mixing means "asymptotically independent", which is the statistical independence(for example, total variation between their joint distribution) goes to 0 as $|t_1 - t_2| \to \infty$. (12). I claim that mixing implies ergodicity and everything we have about the ergodicity applies to mixing process.

Suppose $\{X_t\}$ is a stationary Markov Process, with stationary distribution Q. Denote $L^2(Q)$ the space of Borel-meansurable functions that are square-integrable w.r.t. to measure Q(I will explain meansure later, which is a method to gauge the feasibility of parameters of $\Lambda = (A, b, \pi)$), let

$$\epsilon_t \varphi(x) = E[\varphi(X_t)|X_0 = x]$$

denote the conditional expectation operator on $L^2(Q)$, let

$$Z = \left\{ \varphi \in \text{Ł}^2(Q) : \int \varphi dQ = 0 \right\}$$

denote the space of square-integrable functions with mean 0.

The ρ-mixing coefficients of process $\{X_t\}$ are

$$\rho_t = \sup_{\varphi \in \mathbb{Z}: \|\varphi\|_2 = 1} \|\epsilon_t \varphi\|_2$$

The process is called ρ-mixing if these coefficients converge to zero as $t \to \infty$, and "ρ-mixing with exponential decay rate" if $\rho_t < e^{-\delta t}$ for some $\delta > 0$. For a stationary Markov process, the coefficients ρt may either decay at an exponential rate, or be always equal to one.

The β-mixing coefficients are given by:

$$\beta_t = \int \sup_{0 \leq \varphi \leq 1} \left| \epsilon_t \varphi(x) - \int \varphi dQ \right| dQ$$

The process is called β-mixing if these coefficients converge to zero as $t \to \infty$, it is "β-mixing with exponential decay rate" if $\beta_t < \gamma e^{-\delta t}$ for some $\delta > 0$, and it is "β-mixing with sub-exponential decay rate" if $\beta_t \xi(t) \to 0$ as $t \to \infty$ for some non-increasing function $\xi(t)$ satisfying $t^{-1} \ln \xi(t) \to 0$ as $t \to \infty$.

For an irreducible and aperiodic Markov process, and $0 < rho < 1$, the results

$$\int \|P^n(x, .) - \pi(.)\| \pi dx \leq c\rho^n, \forall n \in Z_+$$

establish the β-mixing property with exponential decay rate of $\{x(t)\}$.

For ρ-mixing, what we can comment on that is: Given (X_t) a strictly stationary Markov process. If $\rho_X(t) \to 0$ as $t \to \infty$ then $\rho_X(t) \to 0$ exponentially fast.

In summary, the following properties holds for the relationships between the HMM model and the ergodicity/β-mixing properties.

Let $\{Y_t\}$ be a generalized hidden Markov model with a hidden chain $\{X_t\}$

(i) If $\{X_t\}$ is geometrically ergodic, then $\{X_t, Y_t\}$ is Markov geometrically ergodic.

(ii) If $\{X_t\}$ is stationary $\beta - mixing$, then $\{Y_t\}$ is stationary β-mixing with a decay rate at least as fast as that of $\{X_t\}$

3.3 Random Coefficient Autoregressive Model

After GARCH model produced, a large portion of GARCH-type model followed up. These models satisfies stationarity and ergodicity, but beyond that, seldom additional properties are explored. For example, we want to know the properties of finite sth moment or judge if they are β-mixing. Those questions, rendered complicated by GARCH type models, are actually holding significant implications on stylized facts such as fat tails and temporal persistency observed in financial data. For nonlinear or non parametric models, they are also regarded as useful tools for constructing varied statistical properties like consistency, asymptotic normality etc.

The occurrence of Random Coefficient Autoregressive Model show a general blue print of solving those complicated problems. Note that GARCH type model is a specific type of generalized RCA model. It's in the vector form

$$\mathbf{X}_{t+1} = \mathbf{A}(e_{t+1})\mathbf{X}_t + \mathbf{B}(e_{t+1}). \ \ t = 0, 1, 2, \ldots,$$

where $\{X_t\}$ is a random R^m-valued process and $\{e_t\}$ is a R^p-valued i.i.d.sequence that satisfy condition e. Note that the support e_t is defined by it's strictly positive density. This means that any e_t that makes the process stated above as zero statistic value will be eliminated and replicated with a process $\{e'_t\}$ that has a positive increasing rate.

Remind that e_{t+1} is independent of the σ field generated by $(X_0, X_1, \ldots, X_t)$. The σ-field on a set or process X means that a field that has a collection Σ of subsets X(includes $\emptyset$), the pair (X, Σ) is called a measurable space. An example is the subset of a $power\ set$: if $X = \{a, b, c, d\}$, one possible σ algebra of X is $\Sigma = \{\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$.

Assumptions of model construction had to be made to cope with the need on investigating financial time series and provide enough efficiency in getting access to those complicated statistical values. I only list those that are distinct.

Denote $\rho(Q)$ denote the largest eigenvalue of matrix Q, which is called the spectral radius. Then $\rho[A(0)] < 1$. Another assumption is that series $\sum_{k=1}^{\infty}[\prod_{j=0}^{k} A(e_{t+1-j})]B(e_{t-k})$ converges almost absolutely. $\prod_{j=0}^{k} A(e_{t+1-j})x$ converges almost surely for any $x \in \mathcal{R}^m$.

Also there exist a positive function V on $\mathcal{R}^m$, a compact set K of $\mathcal{R}^m$ with nonempty interior, and some $\delta > 0$, $\nu > 0$, and $0 < \lambda < 1$ such that

(i) $E[V(X_{t+1})|X_t = x] \leq \lambda V(x) - \nu$ if $s \notin K$;

(ii) $E[V(X_{t+1})|X_t = x] \leq \delta$ if $X \in K$.

Companied with those assumptions, I provide a general concept (if not regarded as a model) that provides estimation of complicated statistical variables. We can therefore design relative mathematical model based on RCA, extract information from this model, and training those models with Baum-Welch algorithm. I will explain this idea clearer in the following section.

4 Application of HMM on some Classic Models in Finance

The idea of constructing Hidden Markov Model based on Stochastic Volatility Model is explicitly explained and summarized in a privileged book ”Inference in Hidden Markov Model” (1).

Here I present some deficiencies of the traditional SV models and show in what way HMM cab resolve these deficiencies.

Let S_k denote the observed price in a financial time series. It's also customary to express the observation as *"relative returns"* $(S_k - S_{k-1})/S_k$ or *"log-returns"* $\log(S_k - S_{k-1})/S_k$, which both present the relative price change over time. Figure 1 show the difference between these expressions.

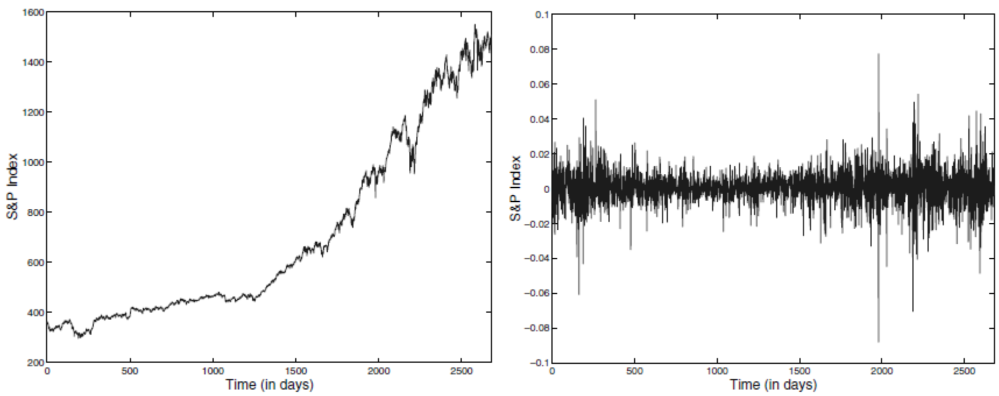


Figure 1: Left: opening values of the Standard and Poors index 500 (S&P 500) over the period January 2, 1990 - August 25, 2000. Right: log-returns of the opening values of the S&P 500, same period. (1)

As I have mentioned before, the principles of designing stochastic volatility models is to consider the correlation between data and the "volatility clustering" phenomenon. In the following subsections I would like to introduce some state-of-art methods

4.1 Sequential Monte Carlo Expectation Maximization (SMCEM)

This method is followed by the generalized linear state-space model in 2.3.. The (log) return of the time series $(S_k - S_{k-1})/S_k$ is generated by weighted mixture of probability densities and each density function corresponds to a hidden state with it's mean and variance. SMCEM help us to derive different volatility levels(states) by deriving the state sequence in the time series.

The estimation procedure consists of three main steps: filtering, smoothing and estimation. Note that we would like to perform model training, and the training prototype is $\{\phi, \tau, \alpha\}$.

The brief introduction of Monte Carlo is: if we want to evaluate $I = \int_{-\infty}^{\infty} g(x)p(x)dx$, we can do it by calculate the mean of the samples $\hat{I} = (1/N)\sum_{1}^{N} g(X_j)$. This methods works well in high-dimensional case, as in here, given certain time, the representation of states vectors are in high dimension (because we are calculating a mixture of samples in multivariate distribution densities).

First we proceed **smoothing step**: Acquire M samples from $f(x_t|Y_t)$(Volatility based on returns) for each t we obtain (3) :

(i) Generate $f_0^{(i)} \sim N(\mu_0, \sigma_0^2)$

For $t = 1, \ldots, n$

(ii) Generate a random number $w_t^i \sim N(0, \tau)$, $j = 1, \ldots, M$

(iii) Compute $p_t^{(i)} = \phi^i f_{t-1}^{(i)} + w_t^{(i)}$

(iv) a. Compute $w_t^{(i)} = p(y_t|p_t^{(i)}) \propto e^{-\dfrac{x_t}{2}} \left(1 + \dfrac{y_t^2 e^{-x_t}}{v-2}\right)^{-\dfrac{v+1}{2}}$

(v) Generate $f_t^{(i)}$ by resampling with weights, $w_t(j)$

This filter is a particle filter, it's used in searching position of some object or estimating the parameter. It's principal is randomly spread "particles" and collect information from each "particle", after applying Bayesian rules we get an more precise range of estimation, we then perform re-sampling based on this estimation.

The second step is **smoothing step**: suppose equally weighted particles $\{f_t^{(i)}\}, i = 1, \cdots, M$ from $f(x_t|Y_t)$ are available for $t = 1, \cdots, n$ from the filtering step

(i) Choose $[s_n^{(i)}] = [f_n^{(j)}]$ with probability $\frac{1}{M}$ For $n-1$ to 0, calculate

$$w_{t|t+1}^{i} \propto exp\left(-\frac{(s_{t+1}^{(i)} - \phi f_t^{(j)})^2}{2\tau}\right)\frac{1}{\sqrt{\pi(v-2)}}\frac{\Gamma\left[\frac{v+1}{2}\right]}{\Gamma\left[\frac{v}{2}\right]} exp\left(1+\frac{y_t^2 e^{-\xi_{t+1}^{(j)}}}{v-2}\right)^{-\frac{v+1}{2}}$$

for each j,

(ii) Choose $[s_t^{(i)}] = [f_t^{(j)}]$ with probability $w_{t|t+1}^{j}$, $(s_{0:n}^{(i)}) = \{s_0^{(i)}, \cdots, s_n^{(i)}\}$ is the random sample from $f(x_0, \cdots, x_n|Y_n)$

(iii) Repeat 1) - iii) for $i = 1, \cdots, M$ and calculate

$$\hat{x}_t^n = \frac{\sum_{i=1}^{M}(s_t^{(i)} - \hat{x}_t^n)(s_{t-1}^{(i)} - \hat{x}_{t-1}^n)}{M}, \hat{p}_t^n = \frac{\sum_{i=1}^{M}(s_t^{(i)} - \hat{x}_t^n)^2}{M}, \hat{p}_{t,t-1}^n = \frac{\sum_{i=1}^{M}(s_t^{(i)} - \hat{x}_t^n)(s_{t-1}^{(i)} - \hat{x}_{t-1}^n)}{M}$$

$$E\left[1+\frac{y_t^2 e^{x_t}}{v-2}\right]^{-\frac{v+1}{2}} = \frac{n(v-2)}{(v+1)\sum_{t=1}^{n} y_t^2 e^{-y_t+v_t}\left[1+\frac{y_t^2 w^{x_t}}{v-2}\right]^{-1}}$$

After we acquire statistics from the distribution, the last step is to train it with Baum-Welch Algorithm:

$$\hat{\phi} = \frac{S_{10}}{S_{00}}, \qquad \hat{\tau} = \frac{1}{n}\left[S_{11} - \frac{S_{10}^2}{S_{00}}\right]$$

$$\hat{\alpha} = \log \frac{n(v-2)}{(v+1)\sum_{t=1}^{n} y_t^2 e^{-y_t+v_t}\left[1+\frac{y_t^2 e^{x_t}}{v-2}\right]^{-1}}\left[n\sum_{t=1}^{n}(y_t - v_t)^{v-1}\right]^{\frac{1}{v-1}}$$

where

$$S_{00} = \sum_{t=1}^{n}(x_{t-1}^n)^2 + p_{t-1}^n,$$

$$S_{11} = \sum_{t=1}^{n}(x_t^n)^2 - p_t^n,$$

$$S_{10} = \sum_{t=1}^{n} x_t^n x_{t-1}^n + p_{t,t-1}^n$$

These formula are vital steps in training the model parameters. It worth to write down for later reference. The simulation process can be estimated by the Jarque-Bera statistics. Since it's not relevant to HMM model, I will omit the details here.

4.2 Application on Option Pricing

Black-Scholes Model, first published in (1997) is a prevalent option pricing model that estimates the price variation over time. This model can be written as:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

where

S, be the price of the stock, which will sometimes be a random variable and other times a constant (context should make this clear);
$V(S,t)$, the price of a derivative as a function of time and stock price;
$C(S,t)$ the price of a European call option and $P(S,t)$ the price of a European put option;
K, the strike price of the option;
r, the annualized risk-free interest rate, continuously compounded (the force of interest);
μ, the drift rate of S, annualized;
σ, the standard deviation of the stock's returns; this is the square root of the quadratic variation of the stock's log price process;
t, a time in years; we generally use: now $= 0$, expiry $= T$;
Π, the value of a portfolio.

The key financial insight behind the equation is that one can perfectly hedge the option by buying and selling the underlying asset in just the right way and consequently "eliminate risk". (cite from wiki).

The famous Black-Scholes model, which is a continuous-time model and postulates a geometric Brownian motion for the price process, corresponds to log-returns that are i.i.d. and with a Gaussian $N(\mu, \sigma^2)$ distribution, where σ is the volatility (the word volatility is the word used in econometrics for standard deviation). The Black and Scholes option pricing model provides the foundation for the modern theory of option valuation. (1)

The traditional Black-Sholes model has deficiencies, however. One of it is that the distribution of the returns usually have heavier tails than the normal distribution, i.e. $p(t)$ is greater when t is (let's say) 3σ farther from the μ of the distribution. This is indicated by the figure 2.

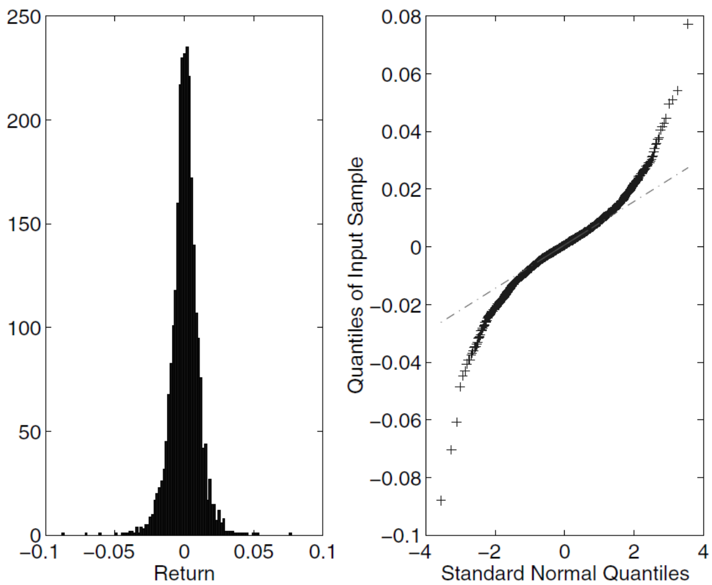


Figure 2: Left: histogram of $S\&P$ 500 log-returns. Right: quantile-quantile regression plot of empirical quantiles of $S\&P$ 500 log-returns against quantiles of the standard normal distribution. (1)

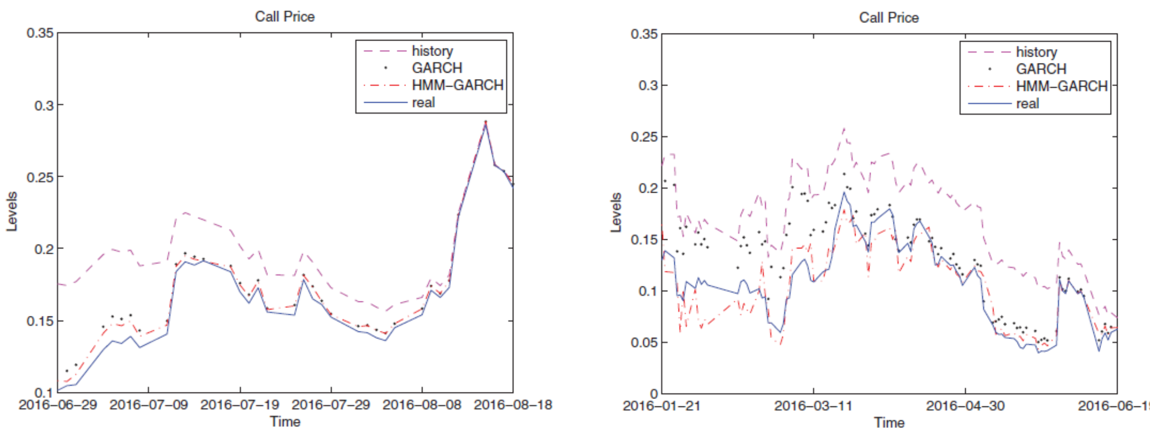
Regard Section 2.3. the generalized volatility model is a good fit for heavy tails. Even with $\varphi = 0$, the model is a Gaussian scale mixture that will give rise to excess kurtosis in the marginal distribution of the data (Thinking Kurtoisis as a measure of the "tailedness"). (1).

Similarly, an experimental algorithm was simulated by (Tang, 2017) et. al. The underlying theorem is also the Bawm-Welch Algorithm.

The process is similar to SMCEM – model the return by the mixture of probabilities with each density function related to the hidden state with mean and variance.

I would like to post their simulation results on multi-option analysis with difference duration and different strike prices.

Figure 3 shows the behaviors of HMM-SV models when strike prices are on RMB 2.05. The simulation result shows that the HMM garch performs better than pure GARCH compared regardless of the option duration. The simulation result shows that the shorter the duration(the earlier the expiration date) is, the smaller the forecast error is .



(a) 2 months (b) 5 months

Figure 3: Call Option (K RMB = 2.05) (4)

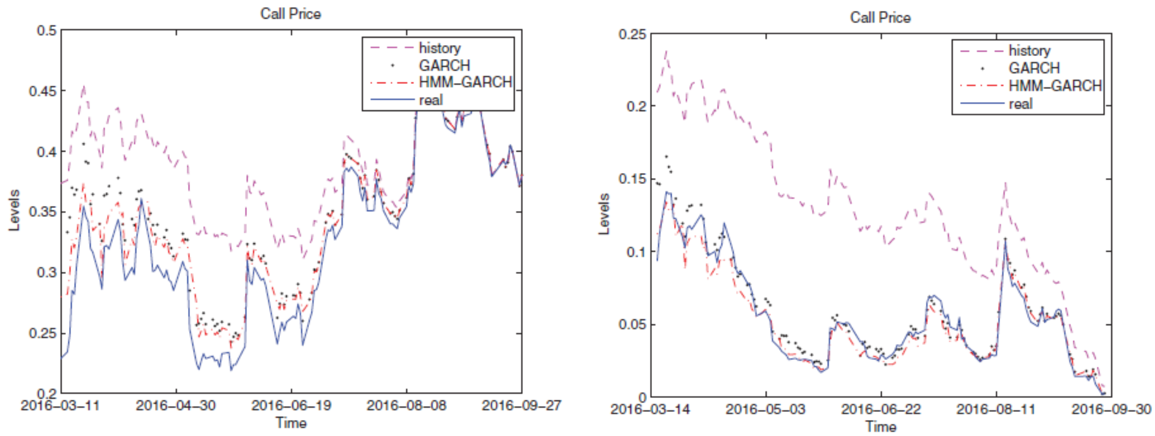
Figure 4 shows on price RMB 2.25 the pricing advantage on RMB 2.25 is not so obvious, it may related with duration and average trading amount.

4.3 Mean Reverting Hidden Diffusion

Previous research focus on the observations where sampling rate is fixed, however with HMM model, sampling rate can also regard as a variable for model estimation. Here I will use the mathematical principles of ergodicity and mixing properties. The point here is that the finite state-space hidden chain in HMM actually present those properties, in the following manner: for $i \geqslant 1$,

$$Z_i = \frac{1}{\sqrt{\Delta}} \int_{(i-1)\Delta}^{i\Delta} \sigma_s dB_s,$$

$$U_i = (\bar{V}_i, V_{i\Delta}), \qquad with\ \ \bar{V}_i = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} V_s ds.$$



(a) $K =$¥1.85 (b) $K =$¥2.25

Figure 4: Call Option (Maturing in Sep., 2016) (4)

We will prove that (Z_i) is an HMM with hidden chain (U_i). Here Δ is some fixed sampling interval. V_i is the volatility (variance) of a process, if $V_t = \sigma_t^2$, then we have

$$d(V_t) = b(V_t)dt + a(V_t)dW_t, \qquad V_0 = \eta$$

This model forms a classical estimation of the unknown parameters of a stochastic volatility model herein. To be more specific, note that $Z_i|V_t, t \geqslant 0 \sim \mathcal{N}(0, \bar{V}_i)$, but the joint distribution of $\bar{V}_i$ and the transition probability of (U_i) remains unknown.

The mean reverting hidden diffusion(volatility) is an adaptation of space model:

$$\begin{aligned} dY_t =& \sigma_t dB_t & Y_0 = 0, \\ dV_t =& \alpha(\beta - V_t)dt + \alpha(V_t)dW_t & V_0 = \eta, V_t = \sigma_t^2 \end{aligned}$$

where $\alpha > 0, \beta > 0$ and $a(V_t)$ may also depend on unknown parameters. Due to the mean reverting drift of (V_t), these models possess some special features.

Before example illustration an proposition is that: Assume that the above hidden diffusion (Vt) satisfies Assumptions in RCA(Section 3.3) and that EV_0^2 is finite. Then, $E\bar{V}_1 = EV_0 = \beta$, and

$$E\bar{V}_1^2 = \beta^2 + var(V_0)\frac{2(\alpha\Delta - 1 + e^{-\alpha\Delta})}{\alpha^2\Delta^2} \quad E\bar{V}_1\bar{V}_2 = \beta^2 + var(V_0)\frac{(1 - e^{-\alpha\Delta})^2}{\alpha^2\Delta^2}$$

One of the interesting model is that: if we set

$$s(v) = Kv^{a-1}exp(\frac{\mu}{v}), \quad M = \frac{1}{kc^2}\frac{\Gamma(a)}{\mu^a}$$

where $\Gamma(a)$ is a usual gamma function, there fore the stationary distribution π has

$$\pi v = \frac{\mu^a}{\Gamma(a)}v^{-a-1}exp(-\frac{\mu}{v})\{v > 0\}$$

which is the inverse gamma function with parameters (a, μ). The moments of π is given by

$$m(p) = E(\eta^p) = \mu^p\frac{\Gamma(a-p)}{\Gamma(a)} \;\; if\ p < a$$

Inverting the formulae leads to consistent estimators of α, β and c^2.

To summarize, we have derived different formula for Hidden Markov Model training algorithms. Those algorithms include the estimation of model parameters and model structure learning.

5 Conclusion

This paper provides a general research on thoughts in designing Stochastic Volatility models from historical process and mathematical derivations. Then I introduced some state-of-arts Hidden Markov solutions to train the SV model and improve the accuracy of model estimation and predictions. First we introduced some basic financial time series model with variant volatility(ARMA) and invariant volatility(GARCH), and derived a general form of the stochastic volatility models. From those models, we want to estimate more statistics by HMM, so we estimate some useful tools for model estimation - ergodicity and β-mixing. To synthesize these qualities into a model we derive the RCA models and made several assumptions for the validity of the model. Last I explore more applications of HMM-SV model as well as some new adaption of HMM-SV model. Those models includes Sequential Monte Carlo Expectation Maximization(SMCEM), option pricing model, and mean reverting hidden diffusion models. This work is a general view with many mathematical and pattern recognition knowledge. I would like to do some simulation on those models to validate their real functional abilities in the future. The research experience will prepared me well in future work of mathematical modeling quantitative researches.

References

1. Capp, Olivier, Eric Moulines, and Tobias Rydn. *"Inference in hidden markov models."* In Proceedings of EUSFLAT Conference, pp. **14-16.** 2009.

2. Carrasco, Marine, and Xiaohong Chen. *"Mixing and moment properties of various GARCH and stochastic volatility models."* Econometric Theory **18**, no. **1** (2002): 17-39.

3. Nkemnole, E. B., and J. T. Wulu. *"Modeling of stock indices with HMM-SV models."* Theoretical & Applied Economics **24**, no. **2** (2017).

4. Tang, Linger, and Xundi Diao. *"Option pricing based on HMM and GARCH model."* In Control And Decision Conference (CCDC), 2017 29th Chinese, pp. 3363-3368. IEEE, 2017.

5. Lewis, Alan L. *Option Valuation Under Stochastic Volatility* Ii. Finance Press, Newport Beach, CA, 2009.

6. Genon-Catalot, Valentine, Thierry Jeantheau, and Catherine Lardo. *"Stochastic volatility models as hidden Markov models and statistical applications."* Bernoulli 6, no. 6 (2000): 1051-1079.

7. Wikipedia `https://en.wikipedia.org/wiki/Financial_signal_processing`

8. J. Gatheral, *The Volatility Surface: A Practitioner's Guide* (John Wiley & Sons, Inc, Indianapolis, IN, 2006).

9. S. Tsay, *Analysis of Financial Time Series* (John Wiley & Sons, Inc, New York, 2002).

10. T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity* (University of California at San Diego, La Jolla and Institute of Economics, University of Aarhus, Denmark, 1986).

11. L. Bergomi, *Stochastic Volatility Modeling* (CRC Press, Taylor & Francis Group, Oxfordshire, UK, 2016).

12. CMU Statistics Course Note `www.stat.cmu.edu/~cshalizi/754/2006/notes/lecture-27.pdf`