A HMM Based on Daily Rate of Return for SSE

Chen Zhiyun 1501213406

Background-Markov Process

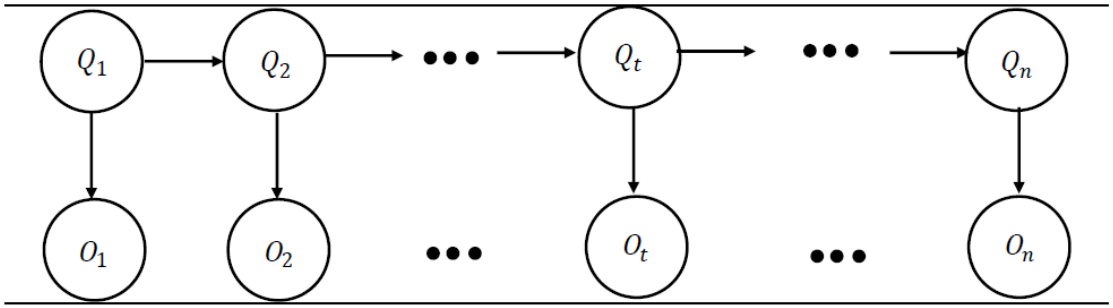
- If a stochastic process has its **future state only related to the present state** and has nothing to do with the state of the past, then we call it Markovy. A stochastic process which have markovy is called Markov process. It is customary to use the conditional probability distribution under continuous time to define the Markov process, where we introduce discrete time and discrete state as an example. If the time and state are discrete, we call the Markov chain.

- A system has a state of S_1, S_2, ..., S_N, with the state of time shift occurred. Assume that the system is in the state of the moment. If the state of the system is only in relation to its state at time t-1, the process of its state change over time is a Markov process. Further, if the state sequence is Q_1, Q_2, ..., Q_t, ..., Q_n, then

$$P(Q_t = S_i | Q_{t-1}, Q_{t-2}, \dots, Q_1) = P(Q_t = S_i | Q_{t-1})$$

Background-Hidden Markov Model

- The variables in the hidden Markov model can be divided into two groups. The first group is the **state variable {$Q1$, $Q2$, ..., Qt, ..., Qn},** Qt said the time of the system state, the length is n, where each state quantity range is {$S1$, $S2$, ..., SN }. It is often assumed that the state variables are hidden and unobservable, so the state variable is also called the hidden variable. The second group is the **observation variable {$O1$, $O2$, ..., Ot, ..., On},** Ot that t time observation value, the value of each observation variable can be discrete or continuous type. For the convenience of discussion, we take discrete observation variables as an example, and assume that the range of values is {$o1$, $o2$, ..., oM}. From the arrows in Figure 1 we can see the dependencies between variables. At any one time, the value of the observed variable depends only on the state variable, regardless of the value of the other state variables and the observed variables. At the same time, the state variable is dependent on the state t-1, which is independent of the previous state. This is the above-mentioned Markov process.



Prediction and Hypothesis

- Hidden Markov model was first applied to speech recognition, and then in other areas have been widely used, such as natural language processing, biological gene sequence analysis and face recognition and so on. So **how can the hidden Markov model be applied to the stock market's forecast, or is it predictive**? Here we need to return to the hidden basic Markovian model of the two basic assumptions, the Markovian and the observation of the independence of the distribution.

- First, where the Markovian refers to the first-order Markovian, the flaw of this hypothesis is that **there is still no more complex correlation between the hidden state**. The first order is the simplest case, so we will start from this model, after all, the first-order hidden Markov model of the timing effect is good? If not, we can further improve and study higher-order, more complex models. Our goal is to look for a Markov model that can better characterize the state of the market.

- Second, **each observation variable is only determined by the hidden state**, this assumption is acceptable. For example, there is a significant difference between the yield of the bull market and the bear market, and the volume of the shock bear is far from the turnover of the fast bull, so we have reason to believe that these two indicators are different in different market conditions. And this distribution only depends on the market state itself.

Data

- From: Wind
- Open, close, high, low of Shanghai Stock composite index and other index such as VMA, RSI, PVT..
- Frequency: daily
- Time period: Jan 4, 2000 to Dec 31, 2016

Empirical Methodology

- Step1. Choose the number of hidden states.
- Step2. Select a single technical indicator as a model input variable.
- Step3. Determine the hidden state of the sample 1~n corresponding to the hidden state category {up, down, shock}, and generate trading signals. In the beginning, for each number i ∈ {1, 2, ⋯, n}, calculate the corresponding long strategy. Then, the sort long strategy in accordance with the yield from high to low and divided into three groups, respectively, on behalf of "up", "shock" and "fall" state set. Finally, replace all numbers with numbers 1,0 and -1, where 1,0 and -1 represent the categories in which the numbers belong, "up", "shock" and "down" to form the final trade signal.

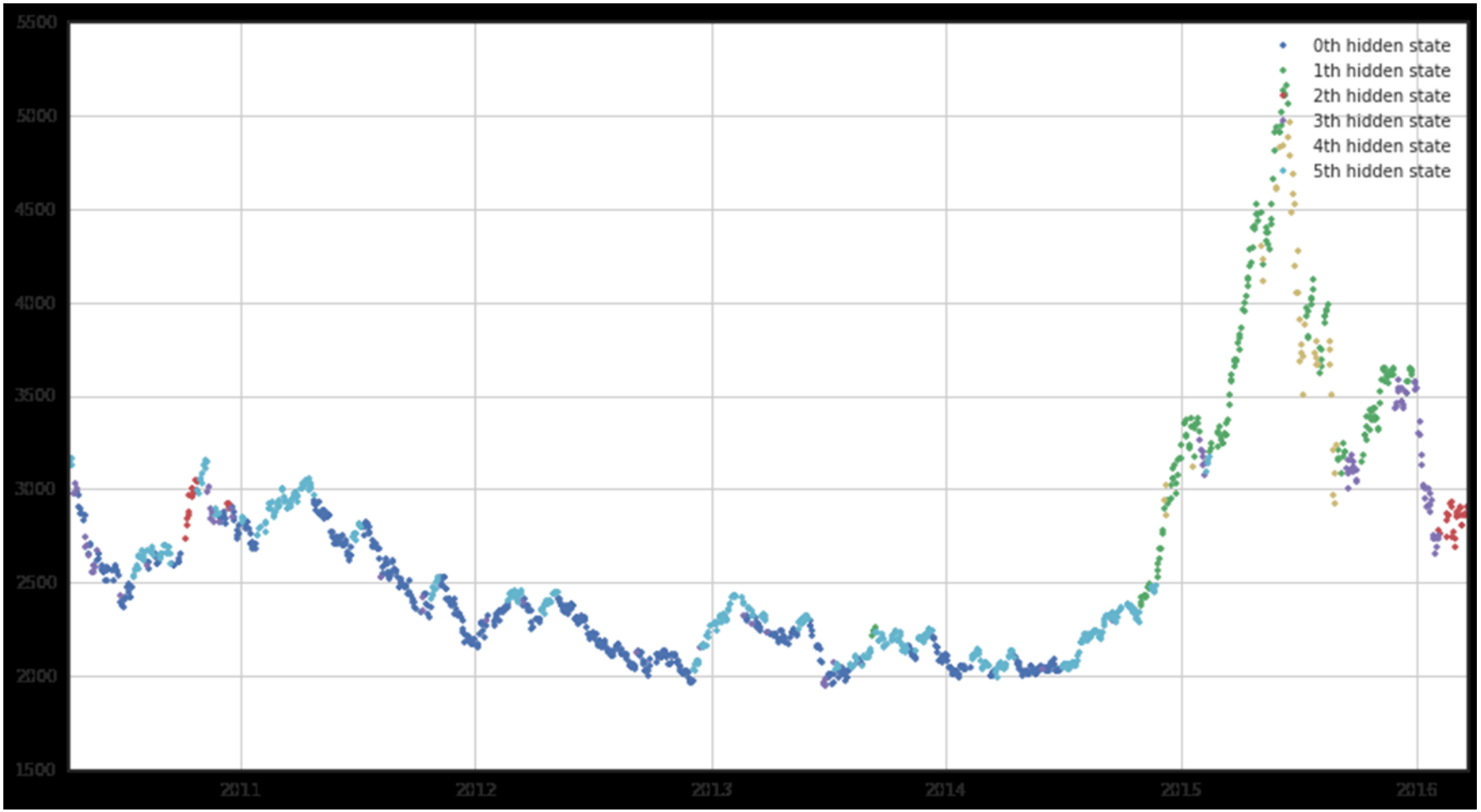
Empirical Methodology

- Step4. Calculate the evaluation index in the sample of the strategy, such as the winning percentage, the transaction frequency, the annualized rate of return, the maximum retracement, the profit taking ratio, the sharp ratio.

- Step5. Repeat Step2 ~ Step4, select the yield retracement ratio and the highest winning percentage of technical indicators.

- Step 6. Use the identified technical indicators for sample detection.

Previous Work in Uqer.io



Thank you