# The Recognition of Investor's Sentiment and the Trading Strategy Based on HMM

**Juan Cheng[1], Chenghu Ma[2], Zhibai Wang[3]**

[1]Xianda College of Economics and Humanities, Shanghai International Studies University

[2]School of Management, Fudan University

[3]PayPal Holdings

1110210@xdsisu.edu.cn, machenghu@fudan.edu.cn, barrywzb@foxmail.com

**Keywords:** Hidden Markov Model, investor's sentiment, transition probability matrix, hidden sequence, observation sequence, timing strategy, back testing

**Abstract:** It is widely believed that the investor's sentiment in securities market is unobservable. All participants can't ignore investor's sentiment since it has great influence on the market. The paper used Hidden Markov Model (HMM) to recognize investor's sentiment in China's A-share market. The investor's sentiment is considered as hidden state in HMM. Since HMM doesn't give the number of the hidden states, the paper sets the number according to Bayesian Information Criterion (BIC). The opening price, the closing price, the highest price, the lowest price and the volume of Shanghai-Shenzhen 300 index (CSI 300) are the observable sequences to figure out the optimal number of the hidden states. The result shows that the number of the hidden states varies from ten to twenty in different periods.In order to evaluate the model, we developed a strategy and did back testing using the historical data of ETF from January 2014 to December 2017.The result shows that the accumulative return of the strategy is greater than that of ETF in most of time except extreme market situation.

## 1. Introduction

Securities market is affected easily by the real economy, macroeconomic policies and international financial environment as well as the investor's sentiment.In the short term, this effect is more obvious. Optimistic investor's sentiment may push the market up, whereas pessimistic investor's sentiment may speed up the drop.

The circuit breakers happened on January 4$^{TH}$ 2016 and January 7$^{th}$ 2016 in China A-share market confirmed the effect fully.

The research about the investor's sentiment has lasted for a long time. The C-CAPM model suggested by Lucas(Lucas,1978) and Breeden(Breeden,1979) relies on the preference of the investor's expected utility and the macroeconomic fundamentals. Preference of the investor's expected utility reflects the investor's sentiment. Epstein and Zin(Epstein and Zin,1989)established general equilibrium capital asset model which relies on two factors —consumption increase rate and market portfolio return. The model makes distinction between risk aversion preference and time preference for inter-temporal consumption. These models rely on how to set the special parameters of the utility function of representative investors. In fact, investor's sentiment is hard to be depicted by a few unknown parameters.

Many researchers tried to seek some trading indicators to measure investor's sentiment. Fisher and Stateman(Fisher and Stateman,2000) found that consumer confident index as an indicator of investor's sentiment could predict the returns of stocks. Baker and Wurgler( Baker and Wurgler,2007) used the principal components abstracted from fund discount rate, market turnover rate, the number of the new issue holders, the average rate of return on the first day of the issue, the bonus rise index to measure investor's sentiment.

In recent years, digging the investor's sentiment from big data has become popular with the

  DOI: 10.25236/icbdai.2018.025

development of behavior finance and the technology of big data analysis, Network language sentiment analysis. Da, Engelberg and Gao( Da, Engelberg and Gao,2015)applied the data which was browsed most often to reveal the change of the market. The data was involving with the topic such as economic recession, unemployment and bankruptcy. They developed a financial climate index to reflect investor's sentiment. And the index had predicted the market reversal in a short term and temporary increase in volatility successfully. Xuejing Meng and Xianglan Meng( Xuejing Meng and Xianglan Meng,2016) used both textual analysis technology and the keywords recommendation system from Baidu to analyzed Micro-blog topics . A sentiment lexicon which could best reflect the investor's sentiment in China's market was obtained from the analysis. They also set up the investor's sentiment index in China securities market based on the sentiment lexicon.

Inevitably, there are some subjective factors in these methods, especially in the measurement of sentiment intensity and the designing of sentiment index.

Strictly, the investor's sentiment, as the aggregation of many individual sentiment is unobservable.

Hidden Markov Model and its algorithms were proposed by L.E. Baum during seventies of twentieth century. Then the model was widely used in speech recognition and natural language processing (Jelinek, 1998), biological information recognition( Churchill, 1989) and financial field (Hassan M R,2005).

Today, HMM is also used to recognize the investor's sentiment and predict the information states and intensity because of its obvious advantages in recognizing hidden states.

Xiaobin Huang, Chunfeng Wang(Xiaobin Huang and Chunfeng Wang ,2011) made a model to depict the unobservable information states in A-share market and analyzed the dynamic relationship between information states over time. They also estimated the information states and information intensity of SSE(Shanghai Stock Change),composite Index and SSE50 index by Bayes inference and Markov Chain-Monte Carlo Method in August 2010. But they did not figure out how to decide the number of the hidden information states.

Some researchers and investors assume that the number of hidden states of the investor's sentiment is constant. They usually set the sentiment states as three states: rise,fall and steady or as five states : rise quickly, rise, steady, fall, fall quickly. This way of setting the state is too subjective though it is convenient for parameter recognition and state estimation.

This paper adopts Bayesian Information Criterion to determine the number of states of investor's sentiment. We chose the Shanghai-Shenzhen 300 Index ( CSI300 for short) as the objective to observe. The index is believed to represent the whole market well. The opening price, closing price, the highest price, the lowest price and the volume are five observation sequences.

We chose data with different transaction frequencies in different periods to determine the number of states of the investor's sentiment by Bayesian Information Criterion.

The results show that the number of states of the investor's sentiment is over ten and near to 20.

To evaluate the method, we developed a trading strategy which took Exchange Traded Fund of CSI300 as the transaction object. We used the data during the period from 2013 to 2017 to do a back testing. The accumulative return of the back testing is greater than that of Exchange Traded Fund in most of the time except the extreme bull market.

That means Bayesian information criteria is an effective way to determine the number of states of investor's sentiment.

In second section of the paper, we introduced Hidden Markov Model. In the third section, we made the model for investor's sentiment recognition. In the fourth section, we designed the trading strategy and in the last section, we made a brief summary.

## 2. Introduction of HMM

HMM is a special case of Markov Model. The model involves two types of state variables. One is hidden state variable represented by symbol "s". Another is observable variable represented by symbol "y". "s" subjects to Markov progress. The number, the range and the transition probability matrix of s may be unknown and yet to be recognized. The probability distribution of y depends on the hidden state variable. Figure 1 illustrates the principle of the model. The number, the range and the transition probability matrix of the hidden states can be inferred by observing the historical sequences of the observable variables.

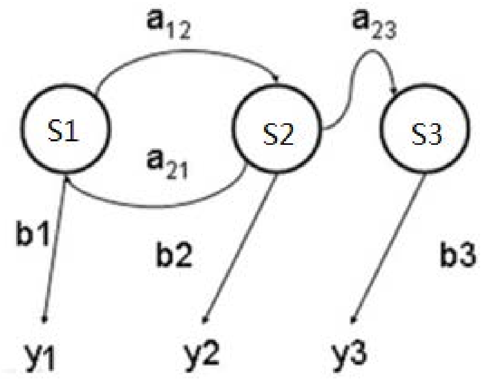


Figure 1 The Principle of HMM.

$s_i$ is the hidden state. $y_i$ is the observation state . $a_{ij}$ is the probability that state $s_i$ transfers to state $s_j$ . $b_i$ is the probability distribution of the observation states under the hidden state $s_i$ .

Suppose $S=\{s_1,s_2,.....s_N\}$ is the range of hidden states and $N$ is the number of hidden states. $Y=\{y_1,y_2,.....y_M\}$ is the range of observation states and M is the number of observation states. $X=(x_1,x_2....x_T)$ is the hidden state sequence. $O=(o_1,o_2......o_T)$ is the correspondent observation state sequence. $A=[a_{ij}]_{N\times N}$ is the transition probability matrix. The element of A is $a_{ij}=P(x_{t+1}=s_j \mid x_t=s_i), i,j=1,2....N$ .

The model assumes that the observation states are independent. That means the observation state depends only on the hidden state at the current moment.

$b_j(k)=p(o_t=y_k \mid x_t=s_j)$ is the probability that the observation state is $y_k$ when the hidden state is $s_j$ at time t. $B=[b_j(k)]_{N\times M}$ is confused matrix which is formed by $b_j(k)$ .

$\pi=[\pi(i)]$,where $\pi(i)=p(x_1=s_i), i=1,2....N$ , is the initial probability vector of hidden states .

HMM is determined by the initial probability vector, transition probability matrix and the confusion matrix. There are three fundamental problems of HMM to solve. The first one is involving the probability of occurrence of a certain observation sequence in the case of given model parameters. Forward-algorithm and Backward algorithm are used to solve the problem. Forward-algorithm defines the forward probability which is the probability of occurrence of a certain observation sequence $o_1,o_2.....o_t$ from original time to t time given that the hidden state is $s_i$ at t time. It is denoted as $\alpha_t(i)=P(o_1,o_2.....o_t \mid x_t=s_i)$ . All forward probabilities can be obtained by recursion formula $\alpha_{t+1}(i)=\sum_{j=1}^{N}\alpha_t(j)\alpha_{ji}b_i(o_{t+1})$ .

Backward algorithm computers the backward probability which is the probability of occurrence of the observation sequence $o_{t+1},o_{t+2}.....o_T$ from time t+1 to T. The probability is denoted as $\beta_t(i)=P(o_{t+1},o_{t+2}.....o_T \mid x_t=s_i,\lambda)$. Similarly, all the backward probabilities can be obtained by recursion.

The second one is involving estimation of the parameters. The maximum likelihood estimation, named as Baum-Welch algorithm is used to find the estimation of A, the transition probability matrix ,and B, the confusion matrix. The algorithm estimates A and B by maximizing the probability of occurrence of the certain observation sequence.

The third one is involving predicting the most likely hidden state sequence through the observation sequence. The problem is also called decoding. Viterbi algorithm is applied to decoding.

### 3. Modeling the investor's sentiment

HMM is used to measure the investor's sentiment for a short term in A- share market. The hidden state sequence is composed of discrete investor's sentiment states. But the number of investor's sentiment states is unknown. The observable sequences are the part of all observable information in the market. The paper firstly set the number of the hidden states since HMM does not tell the number of hidden states or the observable variables.

The paper adopts Shanghai-Shenzhen 300 Index(CSI 300 Index) which was released in 2005 to observe investor's sentiment. The index is a transactional component index. It is believed to reflect the truth of the market since it is weighted by trading volumes of the 50 most influential stocks.

The industries covered by the index are consistent with that of the whole A-share market. So, the index can represent the whole A-share market fully. The investor can trade ETF and corresponding stock index futures after discovering some rules about the index.

This paper chose the opening price, the closing price, the highest price, the lowest price and the trading volume as the observable sequences. To reasonably determine the number of hidden states, we use Bayesian Information Criterions (BIC) to set the number of hidden states. The following formula expresses the algorithm of BIC.

$$BIC = \ln(n) \times k - 2\ln(L)$$

K is the number of the free parameters in the model. n is the number of observed data.

L is the likelihood function, where

$$L = \log(P(s_1)) + \sum_t \log(P(s_{t+1} \mid s_t)) + \sum_t \log(P(Y_t \mid s_t)).$$

The number of the hidden states is best when BIC reaches its lowest value.

Firstly, we selected the data belonging to three typical periods to computer the number of hidden states. These periods were from February to May 2014, January to July 2015 and February to June 2017. In these three periods, the A-share market was in sluggish state, extremely big bull state and rose slowly respectively. Figure 2 shows the value of BIC based on 5-minitues data in each period.

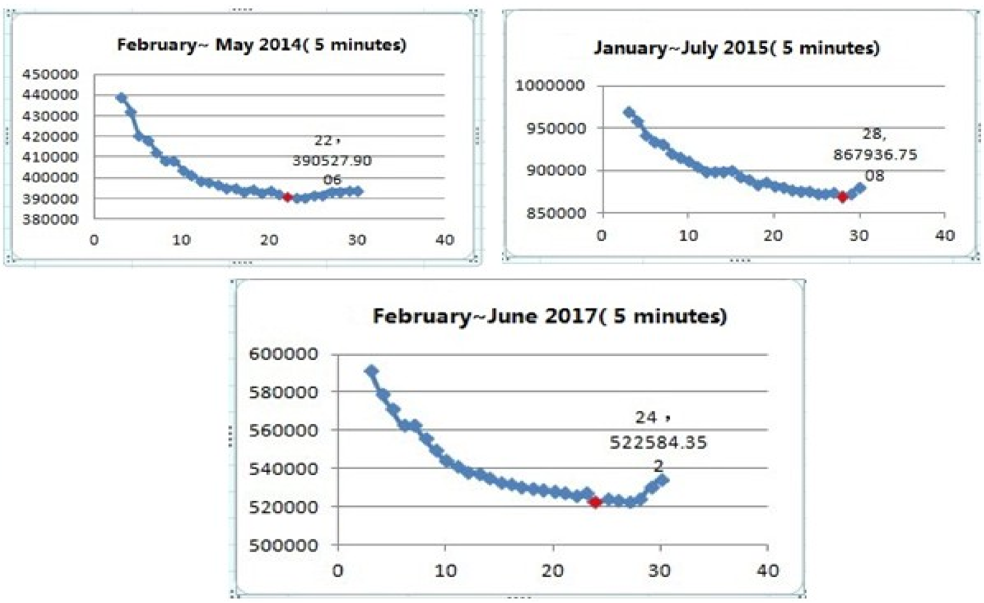


Figure 2.The Value of BIC Based on 5-minitues Data.

The value of BIC first decreases and then increases. In three periods, BIC reaches the minimum when the model had 22,28, 24 hidden states respectively. To further observe the rule, BIC was also computed based on the 60-minites data in each period.The result is shown in Figure 3.

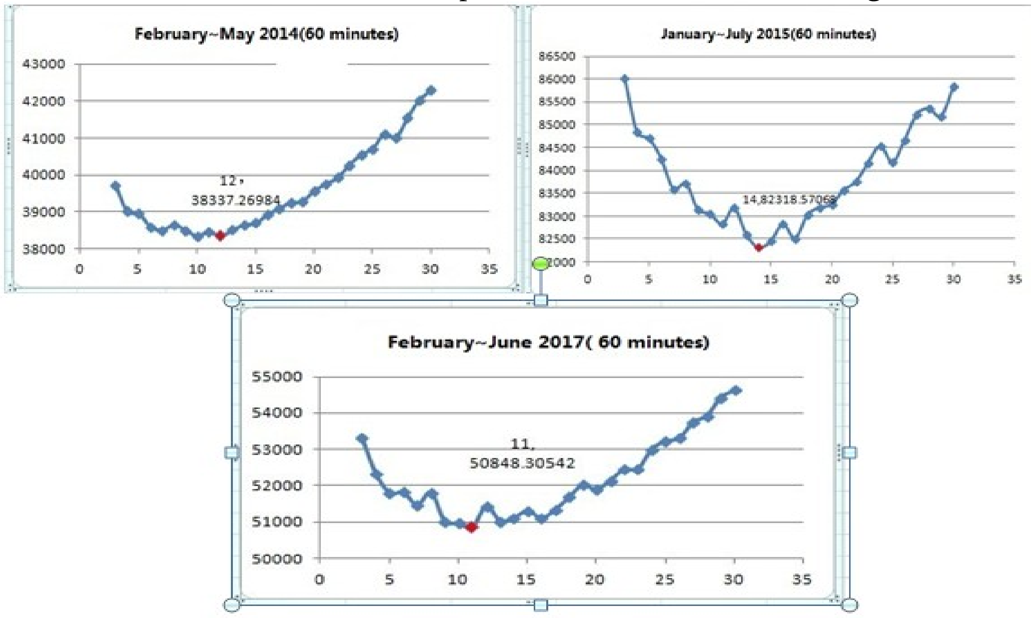


Figure 3 The Value of BIC Based on 60-minites Data

The same rule works with 60-minutes data. The number of hidden states is less than that based on 5-minites. But they are still more than 10. Then, we used daily data to compute BIC in three periods again.

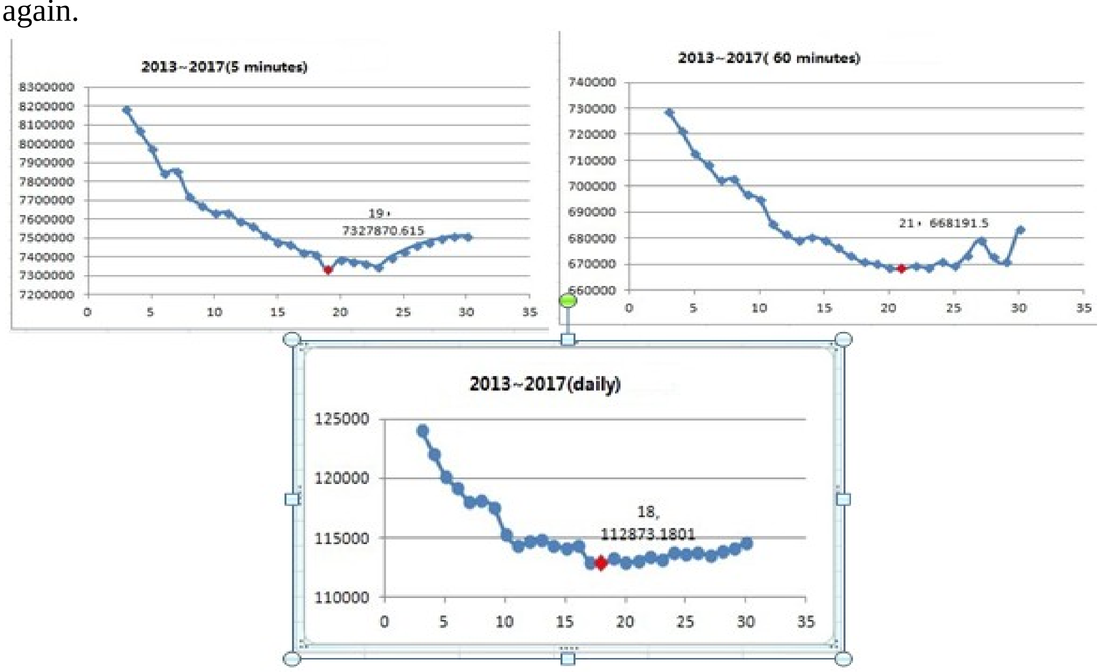


Figure 4 The Value of BIC Based on Daily Data.

The optimal numbers of the hidden states are 19,21 and 18 respectively which are below 20. But in a period up to 5 years, the time of computation was very long when we used 5-minites data.It took about 30 seconds if the number of iterations is 10000.  The time complexity can’t meet the time requirement of quantized transaction.

In terms of investment duration, it is not necessary to compute the value of BIC based on 5-minite data if the investment period exceeds one year. Computing on Daily data is enough.

It can be seen from the above results that 20 hidden states are enough to describe the dimension of hidden state much comprehensively. So, three or five states determined according to subjective experience can’t fully reflect investor’s sentiment states.

For weekly data and monthly data, we suppose there are also such rules. But we didn't compute BIC based on weekly data or monthly data. The reason is that the amount of data is not enough to make the computing converge.  CSI 300 Index has been established only since 2005.

In order to evaluate the method, we extended the research based on the daily data from 2013 to 2017.The optimal number of investor's sentiment state is 18 according to the value of BIC. The transition probability matrix is as following.

The matrix is a diagonally-dominant matrix. That means each state transferred to itself more likely. The states tend to be persistent. Figure 5 shows the hidden states sequence.

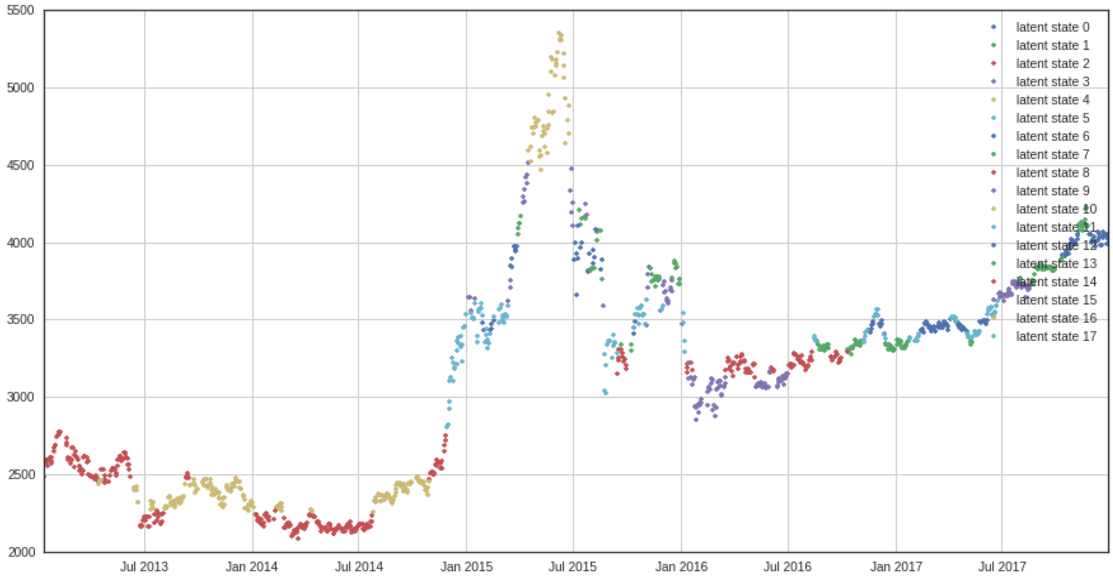


Figure 5 State Recognition Diagram.

Different colors represent different investor's sentiment in the diagram. Figure 6 shows the accumulated return of each investor's sentiment state.

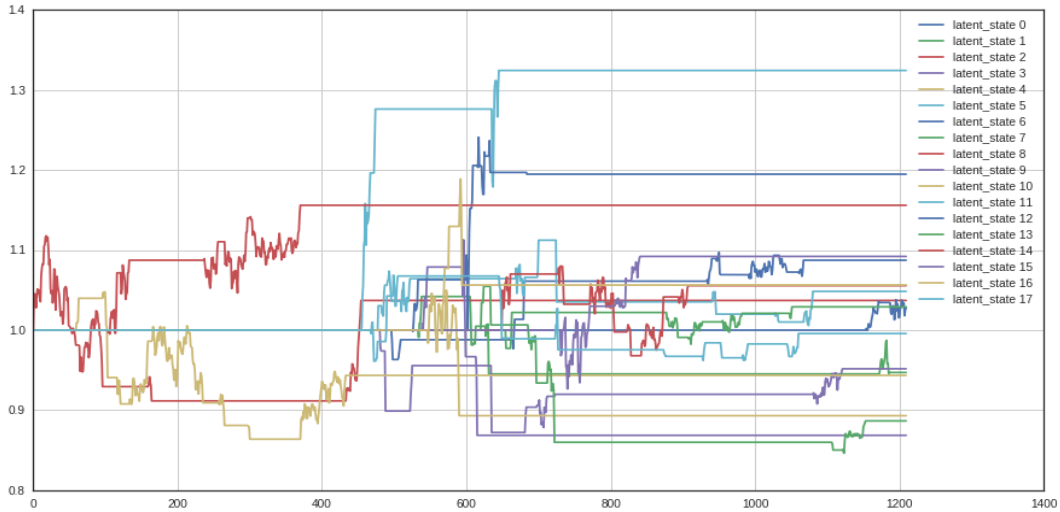


Figure 6 The Accumulated Return.

The horizontal axis is the sequence number of data in chronological order. The vertical axis represents the accumulative return. The horizontal part of each curve indicates that the state didn't appear during that period.

So the accumulative return had no change. "1.0"on the vertical axis means the principle. Exceeding "1.0"means making profit, otherwise means making loss.

We combined state recognition diagram and the accumulative return to identify the state since HMM does not define the meaning of each category.

State 5 had the highest accumulative return which is over 30%. It appeared during January 2015 and July to August 2015.  A- share market reached its peak during June to July 2015. In the half year before June, the stock market was climbing its peak. So the state could be regarded as an extremely optimistic investor's sentiment. It also can be seen in State Recognition Diagram that the state appeared in the first half year of 2015.

State 0 has the accumulative return over 20%. It appeared before and after reaching the peak. But it appeared not often. State 0 mainly appeared in 2017 in which A-share market was slowly going up. So, State 0 can be seen as conservative optimistic sentiment.

State 2, appeared before 2014, had the accumulative return over 10%. In that period, the stock market had tiny volatility. The state is recognized as stable sentiment.

State 4 appeared in peak period. But it had an unsatisfactory accumulative return which was below 10%. When the stock market is at the peak of the bull market, the risk is usually very large and the return may not be ideal.

State 7 has 15% loss. It appeared after July 2015. It was a fall stage after sharp rise. In this market atmosphere, investor is usually pessimistic or even panic.

## 4. The trading strategy

To further evaluate the method, we developed a trading strategy to do back testing during January first 2013 to December 29 2017. Exchange Trading Fund ( ETF for short) of CSI300 was the object since the research was based on the index.

The observable sequences are the opening price, the closing price, the highest price, the lowest price and the volume of the index.

We established HMM based on the data belonging to the previous 250 trading days near the closing time of each trading day. The expected return on the next day under different hidden state was also computed. It is a signal to buy in when the expected return under current hidden state is positive and greater than the median of all expected returns. We hold it until the expected return of the next day is lower than the median of all expected returns.

The commission of ETF is two out of ten thousand and unilateral. The capital gains in the period of short position can make up for the commission. To simplify computing, we ignored both commission and capital gains.

During the whole period, the strategy sent out 37 buying signals. There were 6 signals in 2013. We hold ETF for 159 days according to the first signals on January 11 2013. There was only one signal on October 30 2014 and we hold ETF for 145 days. The buying signals were sent frequently in 2015. But the holding time was short, only about 15 days. The bull market might be the reason. There were five signals in 2016. And the longest holding lasted for 171 days. There were only two signals in 2017 because the market was in downturn.

Since we bought in not often, the capital gains in free time were more than the commission. So, the profit yielded by the strategy was not overestimated. Figure 7 shows the accumulated return of back testing.


Figure 7 The Accumulated Return of Back Testing.

The green dot line denotes the accumulative return of ETF. The blue curve is the accumulative return of the strategy. “1.0”on the vertical axis still means the principal.

It can be seen from Figure 7 that the accumulative return of the strategy reaches 95% with 14.3% annual return.

The return of the strategy is greater than that of ETF except for the period from February 2015 to the end of 2017. In this period, the market was dramatically increasing.

In the period from the middle of 2013 to the end of 2014, the accumulative return curve of the strategy is horizontal. Since the market was falling slowly, it did not trigger the signal of buying in this period.

In the period from the end of 2014 to February 2015, the accumulative return of the strategy was nearly the same with that of ETF.

From the second half year of 2015 to the end of 2017, the accumulative return of the strategy is obviously greater than that of ETF. In this period, the market had been climbing up slowly for a long term.

It is clear that the strategy could perform better in periods increasing or decreasing slowly.

While in the soaring periods, the strategy thinks that investor’s sentiment would be unstable or pessimistic after a quick rise. It is extremely risky to buy in after a quick rise. So the strategy tends to not buy in. That is why the accumulative return is worse than that of ETF in such periods.

Table 2 lists five parameters of the strategy.

Table 2 Parameters of the strategy.

| Volatility | Sharp ratio | Beta Coefficient | Alpha Coefficient | Maximum Drawdown |
|---|---|---|---|---|
| 0.2 | 0.587 | 0.572 | 0.072 | 22.28% |

The accumulative return of ETF was about 71.33%. The volatility of ETF was 0.264。

The strategy has lower risk as well as higher return than those of ETF.

Generally, the strategy can earn stable return in non-extreme market situation. It is also rapid to make transaction decision near the closing time since the calculation speed is quick. In the extreme market situation, the strategy performs worse than ETF. That is an obvious disadvantage of the strategy. In this situation, the strategy is needed to be adjusted according to the market.

## 5. Conclusion

The paper adopted HMM to measure the investor’s sentiment in China’s A-share market. 80% of investors in A-share market are individual investors. They are very emotional. It is difficult to regulate the market. Besides, the capital allocation mechanism can’t work well because of the amount of the irrational investors. Once the crisis breaks out, these investors have weak ability to deal with it. And the impact on the whole financial market is even worse. So, it is important to model and measure investor’s sentiment in China’s A- share market.

HMM can estimate and predict the unobservable Markov sequence. It is suitable for modeling objects that are not directly observed. The algorithms of HMM have been improved by researchers and have high efficiency. We used Shanghai-shenzhen 300 index to make empirical study because the index can represent the whole market very well. The observed variables are the opening price, the closing price, the highest price, the lowest price and the volume of the index. BIC was used to determine the number of the hidden states of HMM. The result shows that the number of the hidden states is about 20 no matter the trading frequency and duration. To assess whether this rule can be applied to investment, we designed a strategy according to the result based on the daily data from 2013 to 2017. The back testing of the strategy shows that the strategy can perform well in most of the time except extreme market situation. Besides, the model based on BIC can identify the change of the hidden state very well and be adaptive for the new hidden states. This illustrates the model has excellent capability in measuring investor’s sentiment of the whole market. It is also possible to design quantitative trading strategies according to the result of the model. The model can be widely used to recognize investor's sentiment and direct investment of each individual share. Certainly, further research is needed on how to improve the return and combining with other quantitative trading thinking. Quantitative strategy traders can also consider a more complex HMM which can measure investor's sentiment more accurately. The investors might gain more return if they can design a better strategy according to a complex HMM.

## References


[1] Baum,L E and Petrie T, 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains, Annuals of Mathematical Statistics. 37(6),pp.1554-1563.

[2] Baum, LE.,Petrie,T. and Sousesl, G, 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, Annuals of Mathematical Statistics. 41(1),pp.164-171.

[3] Brown, G W., Cliff, M T, 2005. Investor Sentiment and Asset Valuation, Journal of Business. 78(2), pp.405-440.

[4] Baker,M., Wurgler, J., 2007. Investor Sentiment in the Stock Market, Journal of Economic Perspectives. 21(2), pp.129-151.

[5] Choi, H.,Varian, H.,2012. Predicting the Present with Google Trends, Economic Record. 88(s1),pp.2-9.

[6] Chunpeng,Yang.,Rengui,Zhang., 2014. Dynamic sentiment asset pricing model, Economic Modelling. 37(2), pp.362-367.

[7] Chenghu,Ma., 2016. The Principle of Financial Economics, Tsinghua University Press. Beijing, 1st Edition.

[8] Da, Z., Engelberg,J., Gao, J., 2015. The Sum of all Fears:Investor Sentiment and Asset Prices, Review of Financial Studies. 28(1),pp.1-25.

[9] Epstein, L., Zin, S.,1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework, Econometrica. 57(4),pp.937-969.

[10] Fama, E F.,,French, K P.,1988. Dividend Yields and Expected Stock Returns, Journal of Financial Economics. 22(1),pp.3-25.

[11] Frederick,Jelinek.,1998. Statistical Methods for Speech Recognition,The MIT Press. London, 2nd Edition.

[12] Fisher, L., Statman, M.,2000. Investor Sentiment and Stock Returns, Financial Analysis Journal. 56(2), pp.16-23.

[13] Forney, J., 2005. The Viterbi Algorithm : A Personal History , Mathematics. 11(2) ,pp. 59-285.

[14] GaryA,Churchill.,1989. Stochastic Models for Heterogeneous DNA Sequences, Bulletin of Mathematical Biology. 51(1),pp.79-94.

[15] Haozheng,Li., Zhiqiang, Liu.,Xianghua, Zhu., 2004. Hidden Markov Models Factored Gaussian Mixtures Densities, The Journal of China Universities of Posts and Telecommunications. 11(3),pp.74-78.

[16] Huichu,Shu.,2010. Investor mood and financial markets, Journal of Economic Behavior &Organization. 76(2),pp.267-282.

[17] Hang,Li., 2012. Statistical Learning Method, Tsinghua University Press. Beijing, 1st Edition.

[18] Jiangde,Yu.,Xiaozhong,Fan.,2007. Application of Hidden Markov Model in Natural Language Processing, Computer Engineering and Design. 28(22), pp.5514-5516.

[19] Jun,Wu., 2016. Beauty of Mathematics, Posts & Telecom Press. Beijing, 2nd Edition.

[20] Kontoyiannis, I., Meyn, S P.,2009. Approximating a diffusion by a finite-state hidden Markov model, Mathematics.127(8),pp.2482-2507.

[21] Lucas, R Jr.,1978. Asset Price in an Exchange Economy, Econometrica. 46(6), pp.1429-1445.

[22] Lee ,W Y., Jiang, C X., Indro, D C.,2002. Stock market volatility excess returns and the role of investor sentiment, Journal of Banking and Finance. 26(12),pp.2277-2299.

[23] Laura,Frieder.,2008. Investor and Price Response to Patterns in Earnings Surprises, Journal of Financial Market. 11(3),pp. 259-283.

[24] Lichi,Yuan., 2008. A Speech Recognition Method Based on Improved Hidden Markov Model, Journal of Central South University (Science and Technology). 39(6),pp. 1303-1307.

[25] Lin,Huang., Chenghu, Ma., Nakata, H.,2017. w-MPS Risk Aversion and the Shadow CAPM: Theory and Empirical Evidence, The European Journal of Finance. 23(11), pp. 947-973.

[26] Pieczynski ,W.,2007. Multisensor triplet Markov chains and theory of evidence, International Journal of Approximate Reasoning. 45(1),pp.1-16.

[27] Qing,Liu., Xiangdong,Yan., 2003. Design and Realization of Genes Recognizing System Based on Hidden Markov Model, Computer Engineering and Application. 40 (24),pp.69-71.

[28] Sharp,W.,1964. Capital Asset Prices: A Theory of Capital Market Equilibrium under Conditions of Risk,. Journal of Finance. 19(3),pp.425-442.

[29] Stein, J., 1996. Rational Capital Budgeting in an irrational world, Journal of Business. 69(4),pp.429-455.

[30] Scott, S L., 2002. Bayesian methods for Hidden Markov models：Recursive Computing in the 21st Century, Journal of the American Statistical Association. 97(457),pp.337-351．

[31] Washha, M., Qaroush,A., Mezghani, M., Sedes,F.,2017. A Topic-Based Hidden Markov Model for Real-Time Spam Tweets Filtering, Procedia Computer Science. 11(2),pp.833-843.

[32] Xiaobin,Huang., Chunfeng, Wang., Zhenming,Fang.,Chunlian,Xiong., 2012. Detecting Chinese Stock Information Based on Hidden Markov Model, Systems Engineering—Theory&Practice. 32(4),pp.713-720.

[33] Xuejing,Meng.,Xianglan,Meng.,Yangyang,Hu.,2016. The Research of the Investor’s Sentiment Based on Text Mining and Baidu Index, Macroconomics. 38(1),pp.144-153.

[34] Yanyi, Xu., Jinhua,Lv., 2016. State Recognition Based on Hidden Markov Model, International Journal of Multimedia and Ubiquitous Engineering. 11(2), pp389-398.

[35] Yongqi, Liu., Lei, Ye., 2018. Monthly streamflow forecasting Based on Hidden Markov Model and Gaussian Mixture Regression, Journal of Hydrology. 561(6),pp.46-159.

Table 1 Transition Probability Matrix.

| | S0 | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S0 | 0.956 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.017 | 0 | 0 | 0 |
| S1 | | 0.958 | 0 | 0 | 0 | 0.0214 | 0 | 0 | 0 | 0 | 0.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S2 | 0 | 0 | 0.771 | 0 | 0.055 | 0 | 0.0174 | | 0.08597 | 0.0343 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.03646 |
| S3 | 0 | 0 | 0 | 0.895 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1053 | 0 | 0 | 0 | 0 | 0 | 0 |
| S4 | 0 | 0 | 0.0585 | 0 | 0.915 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0261 |
| S5 | 0 | 0.042 | 0 | 0 | 0 | 0.796 | 0 | 0 | 0 | 0 | 0.121 | 0 | 0 | 0 | 0 | 0.0405 | 0 | 0 |
| S6 | 0 | 0 | 0.12 | 0 | 0 | 0 | 0.881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.941 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0589 |
| S8 | 0 | 0 | 0.027 | 0 | 0 | 0 | 0 | 0 | 0.9 | 0.072 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S9 | 0 | 0 | 0.037 | 0 | 0 | 0 | 0 | 0 | 0.033 | 0.879 | 0 | 0 | 0 | 0 | 0 | 0 | 0.051 | 0 |
| S10 | 0 | 0 | 0 | 0 | 0 | 0.12 | 0 | 0 | 0 | 0 | 0.759 | 0 | 0.12 | 0 | 0 | 0 | 0 | 0 |
| S11 | 0 | 0 | 0 | 0.094 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.764 | 0 | 0 | 0 | 0.1413 | 0 | 0 |
| S12 | 0 | 0.038 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.9045 | 0 | 0 | 0 | 0.0574 | 0 |
| S13 | 0.034 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.966 | 0 | 0 | 0 | 0 |
| S14 | 0.028 | 0 | 0 | 0 | 0 | 0 | 0.00859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.963 | 0 | 0 | 0 |
| S15 | 0 | 0 | 0 | 0 | 0 | 0.0784 | 0 | 0 | 0 | 0 | 0 | 0.2335 | 0 | 0 | 0 | 0.688 | 0 | 0 |
| S16 | 0 | 0 | 0.013375 | 0 | 0 | 0 | 0 | 0 | 0 | 0.027 | 0.02655 | 0 | 0.0265 | 0 | 0 | 0 | 0.906 | 0 |
| S17 | 0 | 0 | 0.011 | 0 | 0.028 | 0 | 0 | 0.039 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.922 |