A New Investor Sentiment Index Model and Its Application in Stock Price Prediction and Systematic Risk Estimation of Bull and Bear Market

Qiansheng Zhang¹, Sichuang Hu², Libo Chen², Ruixi Lin², Wan Zhang², Ruiying Shi²

¹School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou, China
²School of Finance, Guangdong University of Foreign Studies, Guangzhou, China

Email address: zhouqianh01@126.com (Qiansheng Zhang)

To cite this article:

Received: October 4, 2018; Accepted: December 17, 2018; Published: March 15, 2019

Abstract: Many studies in recent years have shown that investor sentiment affects investor decision-making, which in turn affects stock market volatility and the direction of stock market prices. Since behavioral finance researchers find that linear combinations of stock turnover and popularity indices can greatly reflect stock investor sentiment, this paper aims to construct a new investor sentiment index that can be reasonably applied to predict stock market risk by selecting rational factors. A new investor sentiment index model is first proposed by combining specific monthly new account ratio (SNIA), monthly turnover rate (TOR), popularity index AR, delayed yield (DY) and using principal component analysis approach. Secondly, the indicator is statistically tested. The results of the correlation analysis show that the investor sentiment index is positively correlated with the monthly rate of return, and the result of causal analysis reveals that the investor sentiment index is the Granger cause of the change in yield. Thirdly, a new method is designed to predict the stock price trend by using the presented investor sentiment index. Finally, based on VaR and CoVaR model the investor sentiment index can be utilized to forecast and estimate of systematic risk in the bull or bear market.

Keywords: Investor Sentiment Index, Principal Component Analysis, Prediction of Stock Price, Systematic Risk, Condition at Risk

1. Introduction

On October 9th, 2017, Nobel economics prize was awarded to Richard H. Thaler to praise his contribution in behavioral finance [1]. Many studies in behavioral finance show that investor sentiment affects investors' decision, which continually affects the stock market volatility and the stock market price trend [2]. In recent years, behavioral finance researchers found that there is a close relationship between the investor's sentiment and some stock market indicators [3]. So far, many people have studied the investor sentiment and selected different stock market indicators to construct investor's sentiment index model. For example, Swaminathan [4], Pontiff [5], Huang Shaoan and Liu Da [6] concluded that the discount of closed-end funds can be regarded as one of the indicators of investor sentiment. Fisher and Statman used the percentage of survey members from the Individual Investor Association of America who were optimistic about investing as their sentiment indicators. Through regression tests, they found that for every 1% increase in the index, the future yield of S & P 500 will increase by 0.1% on average. So they used Consumer Confidence Index as an investor sentiment index. In addition, Qiu and Welch Brown [7] also used the consumer confidence index as an indicator of investor sentiment. Baker and Stein [8] represented investor sentiment with average market turnover, and Dong Caiting [9] used turnover as an alternative measure of investor sentiment. Neal [10] used mutual fund net buying as an investor sentiment indicator. By now, the investor sentiment index research is still not perfect and can not reach a unified framework. Moreover, F. A. de Oliveira [11], Ming Zhu [12], Shekhar Gupta [13], Y. Fang

Most of the empirical analysis of investor’s sentiment index model is only limited to the Shanghai A shares [16] or the Shenzhen Stocks A shares [17], and the related papers only select a few stocks on Shanghai Stock Exchange or Shenzhen Stock Exchange to establish a combination of sentiment index model [18]. Therefore, an enhanced investor’s sentiment index model is proposed and extended to the entire Chinese A-share market. And the CSI 300 Index in entire A-share market has become the basic barometer that reflects the overall trend of Chinese stock market. To obtain the investor’s sentiment index model, we employ some methods of principal component analysis in Qingqing Wang [19], Boping Tian [20], Yingmei Zhang [21], Jianjun Li [22]. And the closed-end fund discount index can be referred to literatures by C. Lee A [23], Victor Dragota [24], BodurthaJ, E. Kim [25], Marc W. Simpson [26]. For more information about the stock market and investor’s sentiment can be referred to papers by Kenneth L Fisher [27], Chun Wang [28] and Agrawal. G [29], Xiaohui Qu [30] and Yongkai Ma [31].

2. Construction of STAD Investor Sentiment Index Model

2.1. Basic Concept of Sentiment Index Factors

The effective principles for choosing the sentiment index factors are as follows. Firstly, the selected emotional factors should reflect the investor sentiment in the market. Secondly, the choosen sentiment factors should be widely used. Thirdly, the data of selected sentiment index factor must have availability and completeness.

Based on the above principles, six investor sentiment factors including turnover rate, new account opening ratio, new IPO proportion and macroeconomic indicators are selected, which not only have strong correlation with investor’s sentiment, but also have been validated as effective proxy variables of sentiment index in previous studies. In addition, we add the popular index AR and delayed yield as two technical indicators for the proposed sentiment index model.

2.2. Establishment of STAD Investor Sentiment Index Model

(1) Stationary Test [32]. As the principal component analysis of time series must ensure that the time series is stable, we use the following ADF test model to test the stability of time series of sentiment factor.

\[ \Delta y = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \lambda_i \Delta y_{t-i} + u_t \]

\( y \) represents the first-order lagged term of the sentiment factors such as the delaying rate of return, the popular index AR, the monthly turnover ratio, the monthly new IPO ratio, the monthly new account opening ratio and macroeconomic indicators. This ADF model includes intercepts and trend terms, where \( p \) is the number of hysteresis. The above selected six sentiment factors were used to test data stability. It is proved that all the sentiment factors are stable except the macroeconomic sentiment factor. So, the first-order difference method must be done for the data of macroeconomic factor.

Principal component analysis. Since the scale of delayed yield (DY), popularity index AR, monthly turnover rate (TOR), monthly new IPO ratio (NIPO) and macroeconomic indicators affect investor’s sentiment, the synthetic sentiment index should base on the weighted aggregation of these factors. To get the model of investor’s sentiment which has high correlation between monthly rate of return and investor’s sentiment, the above-mentioned six factors are combined in different ways and do principal component analysis respectively.

In the following, the factor’s combination containing delayed yield (DY), popularity index AR, monthly turnover rate (TOR) and specific monthly new account opening ratio (SNIA) is taken as an example to present the process of principal component analysis. Use the princomp function provided in matlab to process the sample matrix combined by DY, AR, TOR and SNIA. The principal component analysis result is listed in the following Table 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Different indexes</th>
<th>Eigenvalue</th>
<th>Difference value</th>
<th>Rate of contribution (%)</th>
<th>Accumulated contribution rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z1</td>
<td></td>
<td>1.8713</td>
<td>0.8590</td>
<td>46.7825</td>
<td>46.7825</td>
</tr>
<tr>
<td>Z2</td>
<td></td>
<td>1.0123</td>
<td>0.2536</td>
<td>25.3083</td>
<td>72.0908</td>
</tr>
<tr>
<td>Z3</td>
<td></td>
<td>0.7587</td>
<td>0.4010</td>
<td>18.9674</td>
<td>91.0582</td>
</tr>
<tr>
<td>Z4</td>
<td></td>
<td>0.3577</td>
<td></td>
<td>8.9418</td>
<td>100.0000</td>
</tr>
</tbody>
</table>

Here are the corresponding eigenvector matrix by the four eigenvalue:

\[
\begin{bmatrix}
0.4854 & -0.1892 & 0.8086 & -0.2733 \\
-0.1034 & 0.9513 & 0.2901 & 0.0161 \\
0.5800 & 0.2254 & -0.4992 & -0.6030 \\
0.6460 & 0.0920 & -0.1130 & 0.7493
\end{bmatrix}
\]
The figures of line 1 to line 4 in the above matrix respectively are the coefficient of each factor in the first principal component to the fourth principal component, so the principal component are displayed in following Table 2.

Finally, using the contribution rate of each principal component we get the investor sentiment index (STAD) as follows.

\[
\text{STAD} = 0.3081 \times \text{DY} + 0.2488 \times \text{AR} + 0.1798 \times \text{TOR} + 0.3711 \times \text{SNIA}
\]

### Table 2. Coefficient matrix of principal components.

<table>
<thead>
<tr>
<th>standardized variable</th>
<th>Different indexes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The first principal component coefficient</td>
</tr>
<tr>
<td>Delayed Yield (DY)</td>
<td>0.4854</td>
</tr>
<tr>
<td>Popularity index AR</td>
<td>-0.1034</td>
</tr>
<tr>
<td>Monthly Turnover Rate (TOR)</td>
<td>0.5800</td>
</tr>
<tr>
<td>Specific Monthly New Account Opening Ratio (SNIA)</td>
<td>0.6460</td>
</tr>
</tbody>
</table>

### 3. Statistical Test and Stock Price Forecasting Ability Test of STAD Investor’s Sentiment Index

#### 3.1. Correlation Coefficient Test of Investor Sentiment Index

First, we can compute the correlation coefficient between investor sentiment and rate of return as following formula,

\[
\rho = \frac{\text{cov(Emotion, Yield)}}{\sqrt{\text{D(Emotion)} \times \text{D(Yield)}}}
\]

where Emotion is time series of monthly investor’s sentiment, Yield is time series of rate of return of Shanghai and Shenzhen 300 index.

PCA is done for all the combination of the above-mentioned factors. The correlation coefficient between Yield and each emotion index based on all the possible factor groups are computed in Table 3.

### Table 3. Correlation coefficient of each factor’s combination.

<table>
<thead>
<tr>
<th>group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho)</td>
<td>0.2001</td>
<td>0.6241</td>
<td>0.0898</td>
<td>0.2349</td>
<td>0.0157</td>
<td>0.1530</td>
<td>-0.1748</td>
</tr>
<tr>
<td>group</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.1184</td>
<td>0.2120</td>
<td>0.1062</td>
<td>0.1982</td>
<td>0.0123</td>
<td>0.6235</td>
<td>0.5596</td>
</tr>
<tr>
<td>group</td>
<td>26</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>45</td>
<td>46</td>
<td>56</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.6191</td>
<td>0.1051</td>
<td>0.0899</td>
<td>0.1128</td>
<td>0.0164</td>
<td>0.1538</td>
<td>0.1397</td>
</tr>
<tr>
<td>group</td>
<td>123</td>
<td>124</td>
<td>125</td>
<td>126</td>
<td>134</td>
<td>135</td>
<td>136</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.6733</td>
<td>0.6639</td>
<td>0.4406</td>
<td>0.5384</td>
<td>0.2195</td>
<td>0.0279</td>
<td>0.1019</td>
</tr>
<tr>
<td>group</td>
<td>145</td>
<td>146</td>
<td>156</td>
<td>234</td>
<td>235</td>
<td>236</td>
<td>245</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.0193</td>
<td>0.1723</td>
<td>0.0682</td>
<td>0.6410</td>
<td>0.5022</td>
<td>0.2279</td>
<td>0.0806</td>
</tr>
<tr>
<td>group</td>
<td>246</td>
<td>256</td>
<td>345</td>
<td>346</td>
<td>356</td>
<td>456</td>
<td>1234</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.5372</td>
<td>0.1267</td>
<td>0.0204</td>
<td>0.1660</td>
<td>0.1621</td>
<td>0.0737</td>
<td>0.6758</td>
</tr>
<tr>
<td>group</td>
<td>1235</td>
<td>1236</td>
<td>1245</td>
<td>1246</td>
<td>1245</td>
<td>1345</td>
<td>1346</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-0.0312</td>
<td>0.6744</td>
<td>0.2286</td>
<td>0.3084</td>
<td>0.2383</td>
<td>0.0312</td>
<td>0.1872</td>
</tr>
<tr>
<td>group</td>
<td>1356</td>
<td>1456</td>
<td>2345</td>
<td>2346</td>
<td>2356</td>
<td>2456</td>
<td>3456</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.0246</td>
<td>0.0460</td>
<td>0.4958</td>
<td>0.6581</td>
<td>0.1416</td>
<td>0.1127</td>
<td>0.1237</td>
</tr>
<tr>
<td>group</td>
<td>12345</td>
<td>12346</td>
<td>12356</td>
<td>12456</td>
<td>12345</td>
<td>123456</td>
<td>123456</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.5423</td>
<td>0.3704</td>
<td>0.3958</td>
<td>0.2127</td>
<td>0.0454</td>
<td>0.0911</td>
<td>0.0911</td>
</tr>
</tbody>
</table>

Remark: ‘1’ represents Delayed Yield (DY), ‘2’ represents Popularity index AR, ‘3’ represents Monthly Turnover Rate (TOR), ‘4’ represents Specific Monthly New Account Opening Ratio (SNIA), ‘5’ represents Monthly New IPO Ratio (NIPO) and Macroeconomic Indicators.

Obviously, the factor group “1234” has the highest correlation 0.6758. So, the factors DY, AR, TOR and SNIA are chosen for establishing the investor’s sentiment index model. In the following graph, both the STAD index and monthly rate of return of CSI300 are drawn, one can see that the correlation between STAD sentiment index and monthly rate of return is very high.

From the figure 1, there are two high level of investor’s sentiment, which is in the 43th month and the 89th month, that is around in December 2008 and January 2013. At the end of year 2008, in order to tackle with the finance crisis the central bank carried out Four Thousand Billion Stimulus Plan and made cuts to interest rates and required reserve ratios to release liquidity. The stock market began to complete the system of security market, build the Growth Enterprises
Market and Share Price Index Futures. With the stock market changing from bear market to bull market the investor’s sentiment and the rate of return of the stock market were very high. At the beginning of the year 2013, the rate of return in this month and the investor’s sentiment were at the highest level though the stock market was in the bear market.

Figure 1. STAD index and monthly rate of return of CSI300.

3.2. Granger Causality Test

In order to prove that the proposed STAD sentiment index is the reason that causes the change of stock yield, we first do the Stationarity Test of STAD index and stock return index to assure the two time series are stationary. Then we make Granger causality test for STAD sentiment index and stock return. The causality test results are displayed in the following Table 4.

<table>
<thead>
<tr>
<th>Original hypothesis</th>
<th>Observed value</th>
<th>F statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock yield is not the Granger cause to STAD</td>
<td>135</td>
<td>1.44203</td>
<td>0.1862</td>
</tr>
<tr>
<td>STAD is not the Granger cause to stock yield</td>
<td>2.36493</td>
<td>0.0213</td>
<td></td>
</tr>
</tbody>
</table>

Remark: Duration of Lag of the causality test is 8.

From Table 4, one can see that STAD sentiment index is exactly the Granger cause of the stock yield, but Stock yield is not the Granger cause of STAD sentiment index.

3.3. Prediction Ability Test

According to the above Granger causality test result, it is supposed that investor’s sentiment can be used to predict the stock price, which means the sentiment index will affect stock price. When STAD sentiment index is high, the overall market sentiment tends to be optimistic and market price will rise, but when the sentiment is low, the whole market sentiment tends to pessimistic and market price will fall. In order to accurately determine the level of STAD sentiment index, the threshold of STAD is set. When STAD sentiment index is above this threshold, the investor’s mood is optimistic. When STAD is lower than this threshold, investor sentiment is pessimistic.

In order to get this specific threshold value, a mathematical experiment is performed using matlab. Assuming STAD is above the threshold of the experiment, the price increase is successful. In addition, assuming STAD is below the threshold for the experiment, the price drop is also successful. One hundred and forty-three of experiments are then performed for different STAD sentiment thresholds. All the experimental results are displayed in the following Figure 2.

From Figure 2, one can see that the optimal threshold value is 0.278, where the probability of predicting successfully reaches up to 0.8531. Based on the Prediction ability test, we can use the STAD sentiment index value to predict the stock price trend successfully with probability 85.31%. As long as the investors acquire rational value of factors DY, AR, TOR and SNIA, the risk of decision error will reduce and the level of yield will increase.
4. Prediction and Estimation of Systematic Risk from STAD Investor’s Sentiment

4.1. The Relationship Between Investor’s Sentiment and Systematic Risk

Investor’s sentiment may enlarge the effect of systematic risk in the extreme market condition. In the bull market, the investor’s sentiment is overheating and the bubble is tremendous, which enlarge the systematic risk. On the contrary, in the bear market investors scramble to offload shares, transaction volume decreases and systematic risk increases. According to this economic phenomena, it can be reasonably supposed that the investor’s sentiment can be used to explain the systematic risk, and this explanation will be strengthened when the market is in the irrational state.

4.2. Measurement of Systematic Risk

To verify the relationship between investor’s sentiment and systematic risk, it is necessary to do empirical analysis. First, we should determine the effective tool to measure the systematic risk. At present, there are a lot of tools to measure the systematic risk, including β coefficient, Value at risk (VaR) [33], Conditional value at risk (CoVaR) [34], SRISK and so on. As β coefficient is not intuitive and lack of timeliness, this article regards VaR and Conditional value at risk as the criterion of measuring systematic risk [35].

VaR refers to the maximum expected loss with a confidence level at a certain amount of time. To obtain VaR formula which has a given confidence level $q=0.05$, suppose that the initial value is $V_0$ and the actual rate of return during the period is $r$, the expected rate of return with confidence level of 0.05 is $r'$, standard deviation of earnings is $\sigma$. Suppose that rate of return obeys normal distribution $N(r, \sigma^2)$, and $u_q$ is $q$ quantile of this normal distribution. So $r' = r - u_q$ and VaR is $V_0(r - u_q)$.

The monthly data of VaR is estimated by using the daily data of the CSI 300 Index to match the STAD sentiment index in the time series.

4.3. Empirical Analysis of Systematic Risk in Stock Market and Investor’s Sentiment

As is well known, the investor’s sentiment index in the irrational state has a fairly strong interpretation of systematic risk. When the stock market enters into bull and bear market, it will meet the irrational condition. Below is the monthly K chart of Shanghai and Shenzhen 300 index from year 2015 to year 2017.
In the Figure 3 of Shanghai and Shenzhen 300 index monthly data, we choose the following four groups as abnormal samples. (1) The index data from October 2006 to October 2007 is increased from 1464.47 to 5688.54, which is regarded as the bull market sample. (2) The index data from November 2007 to October 2008 is decreased from 4737.41 to 1663.66, which is regarded as the bear market sample. (3) The index data from September 2014 to May 2015 is increased from 2450.99 to 4840.83, which is regarded as the bull market sample. (4) The index data from May 2015 to September 2015 is decreased from 4473 to 3202.95, which is regarded as the bull market sample. Additionally, the other samples are all regarded as normal sample.

According to the above result and the optimal threshold value 0.278, it is easy to find that the systematic risks are all very high no matter the STAD sentiment index is higher or lower than the threshold value. Meanwhile, the threshold value can be used not only in predicting the stock price but also in estimating the systematic risk based on the deviation between STAD index and threshold value.

4.4. Empirical Analysis of CoVaR and Investor’s Sentiment in all Industries

To explore the relationship between STAD sentiment index and the systematic risk of Shanghai and Shenzhen 300 index, this paper proposes CoVaR method. Assume CoVaR$_{ij}^q$ (i represents the No. j industry and i represents the system) represents the maximum loss of confidence level q when the loss of Shanghai and Shenzhen 300 index is $x_i = VaR_i^q$. The conditional value at risk can be used to judge the risk spillover effect of Shenzhen and Shanghai 300 index in each industry.

Next, we use the following quantile regression models to calculate CoVaR$_{ij}^q$ and $x_i^{ij}$. 

$$x_i^{ij} = \alpha_i + \beta_i \cdot x_i$$

$$CoVaR_{ij}^q = \alpha_q \cdot VaR_i^q + \beta_q$$

The intercept terms $\beta_q$ and slope items $\alpha_q$ of the above quantile regression equation of different industries are listed in Table 6.
Finally, the condition value at risk of each industry are calculated as in Table 7.

<table>
<thead>
<tr>
<th>Time</th>
<th>08.11-10.03</th>
<th>10.05-11.01</th>
<th>11.02-12.11</th>
<th>13.06-14.04</th>
<th>15.06-17.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAD</td>
<td>0.3818</td>
<td>0.2953</td>
<td>0.2700</td>
<td>0.262</td>
<td>0.292</td>
</tr>
<tr>
<td>Industry with the highest risk</td>
<td>Estate</td>
<td>Medicine</td>
<td>Consumption</td>
<td>Medicine</td>
<td>Estate</td>
</tr>
</tbody>
</table>

From Table 7, one can see that the investor’s sentiment is very huge from Nov 2008 to Mar 2010 in the bull market, the Estate industry has a higher risk contribution to the whole SCI300 index than any other industries. Moreover, the investor’s sentiment is close to 0.3 from May 2010 to Jan 2011 in the normal condition, the medical industry has the highest risk contribution rate. From Feb 2011 to Nov 2012, the investor’s sentiment is lower than 0.3 that is close to the condition of bear market, the Consumptive industry becomes the highest risk contributory industry. Also, the Medical industry has the highest risk contribution from Jun 2013 to Apr 2014 in the bear market. From Jun 2015 to Oct 2017, the Estate industry has the highest risk contribution in the normal condition. In conclusion, Estate industry has the highest risk contribution in the most bull or normal market. And the Medicine industry concentrates in the normal market and bear market, the consumptive industry concentrates in the bear market.

5. Conclusion

This paper constructs a new type of investor’s sentiment index using principal component analysis method, reveals the correlation and causality between the investor’s sentiment index and monthly return rate. Then the applications of the proposed investor’s sentiment index in the prediction of stock price and the explanation of systematic risk are also discussed. Due to that STAD sentiment index is a composite score of the deferred yield, the popularity index AR, the monthly turnover rate and the monthly new account opening ratio in a given month, the investor’s sentiment index model fully considers the trends, historical data and social phenomena influence on investor’s emotion. Additionally, the lag yield and popularity indexes are introduced in the proposed sentiment index model, which are closer to small and medium-sized investors because most small and medium-sized investors have a limited rationality in investment and are easily influenced by historical reasons and trends. Due to that the proportion of retail investors in the Chinese stock market is quite large, the presented STAD sentiment index model specifically considers the emotions of small and medium-sized investors.

Acknowledgements

This paper is supported by the Natural Science Foundation of Guangdong Province, China under Grant 2018A030313996 and 2017A030313435.

References


