

## Investor Sentiment and Option Implied Volatility

Yiwen Zhou<sup>1\*</sup>, Xuanchen Zhang<sup>2</sup>

<sup>1</sup>Faculty of Engineering, The Chinese University of Hongkong, Hongkong, China

<sup>2</sup>King's Business School, King's College London, London, UK

\*Corresponding Author: Yiwen Zhou

**Keywords:** Implied volatility, Investor sentiment, Option pricing, Time-series analysis

**Abstract:** This paper empirically investigates how we can capture sentiments and subjective expectations of investors from Chinese option markets. Through principal component analysis of six representative sentiment factors including closed-end fund discount, share turnover, number of IPOs, average first day return of IPO companies, amount of new investors, and consumer confidence level, we find they are strongly correlated with and have statistically significant predictability of options implied volatilities. The estimations are robust when adding macroeconomic factors as control variables. Our findings have two main contributions. Firstly, we successfully prove the rich investor sentiment information contained from option implied volatility, which can be considered as a market timing device. Secondly, we exploit advanced machine learning methodologies for variable dimensionality reduction and obtain superior results.

### 1. Introduction

Traditional finance theorems like efficient market hypothesis [1], capital asset pricing model [2] and Black-Scholes option pricing formula [3] all retain one core assumption that is market participants are all rational. Investors absorb market news and company announcements quickly and update their beliefs normatively in the way described by Bayes' law, and they trade off risks and returns fundamentally like mathematicians [4]. Even if price deviations occur, it will be correctly captured by sophisticated investors who then react immediately to pull the asset prices back to their fundamental values. Such behaviors are referred to as arbitrage, from which traditional scholars are confident about the efficiency of real markets. While these traditional models are helpful in providing university students with basic views of finance due to their elegance and simplification, they are not consistent with real markets and empirical tests.

One reason for this phenomenon is the mechanism of arbitrage may contain fundamental risk, noise trader risk, liquidation risk and implementation risk when performing in the real practice [4]. Suppose a professional fund manager longs a stock which he believes is undervalued after cautious analysis, it is unknown when the price of stock would move up. The stock may come up with a piece of bad news leading to further divergence and losses. Of course, the manager knows this fundamental risk and he would short a substitute at the same time. The problem is that substitute securities are rarely perfect, making it not possible to remove all the fundamental risks. Noise traders would worsen the mispricing exploited by the fund manager as well. If one has granted the possibility that a price can be different from its fundamental value, then he might also grant the possibility that future price movements will increase in the divergence. Moreover, since the fund manager is just the agent to operate the money, if his arbitrage strategy does not perform well in the short run, investors are highly possible to question his capability and withdraw their investments. Thus, the liquidity risk will result in the manager being not able to insist on his trading strategy. Furthermore, arbitrage strategies are costly in the real markets and sometimes the returns could not even compensate the transaction costs and short-selling constraints.

Based on the limit to arbitrage, it is easier to understand why trading strategies with obvious arbitrage spaces could outperform markets robustly in the long-term. However, the idea of limit to arbitrage is not enough to fully explain various market anomalies, like equity premium puzzle

indicating stock markets earn significantly excessive beta-adjusted returns than bond markets [4]. Therefore, more and more researchers start to criticize the assumption of investor rationality of traditional models and they argue investor sentiments and subsequent irrational behaviors would truly and tremendously influence risky asset returns.

The rest of this paper is organized as follows. Since analysing investor sentiment is the main topic of this paper, we will provide a brief review about this filed in the next section. Section 3 describes the data and Section 4 explains our estimation methodology. Section 5 will demonstrate results of empirical analysis as well as robustness tests. Section 6 claims the implications of our works and Section 7 concludes.

## **2. Literature Review**

The most common irrational behaviors resulting from investor sentiments can be referred to as overreaction and underreaction to the market. Even sophisticated investors may judge asset future performance depended on its historical returns, with previous winners being overvalued and previous losers being undervalued. De Bondt and Thaler (1985) [5] use American stocks data to sort portfolios based on their past returns and argue that past losers significantly outperform past winners which cannot be explained by traditional finance theories. Their empirical findings provide strong support for the idea that investor sentiment does influences subsequent asset returns. It is obvious that investors are willing to pay more (less) to purchase overvalued (undervalued) asset and accept lower (higher) returns in the future, which is described as reversal anomaly. Another famous anomaly is called momentum, meaning past winners can still earn excessively subsequent returns. Barberis, Shleifer, and Vishnu (1998) [6] build a model of investor sentiment to systematically analyse phenomena of long-term momentum and short-term reversal.

With the development of time-series analysis and factor modelling, researchers start to formulate sentiment factors with quantitative data and empirically estimate their explanatory and predictability power of asset returns. Neal and Wheatley (1998) [7] apply the level of discounts on closed-end funds, the ratio of odd-lot sale to purchases and net mutual fund redemptions to build sentiment factors, and Baker and Wurgler (2006) [8] formulate six other sentiment factors to test their implications on stock returns. Apart from the statistically significant results, they also find investors are willing to pay higher prices for lottery-like stocks such as small stocks, young stocks, high volatility stocks, low profitability stocks and distressed stocks, when their sentiments are bullish. Besides stock markets, researchers engage proactively on the investigation of relationship between investor sentiment and option markets as well. Han (2008) [9] exploits S&P 500 options data and indicates option volatility smile is steeper and index returns are more left skewed when market is in the bearish sentiment.

Inspired by these works, we aim to capture investor sentiment information by linking a bridge between stock market and option market. We would like to apply representative sentiment factors built based on stock market data to analyse whether they could explain the time-series variation of option prices. Detecting option prices or option returns directly is not suitable since option contracts have expiration dates and their prices will be influenced heavily by time to maturities. Instead, option implied volatility is considered as a good indicator for monitoring investor sentiments and thus we decide to choose it as the dependent variable of time-series regression. Since the number of explanatory variables is large and they are correlated with may lead to additional model errors, we will use machine learning methods to implement dimensionality reduction and keep the most valuable components.

## **3. Data**

### **3.1 Option Data**

Raw data come from the database of Choice and Tushare which includes all options contract information traded in Shang Stock Exchange. Since most research results in this field are obtained

from US markets, we decide to use data from a different market and test whether results are robust worldwide. However, only 50ETF option has a relatively long trading history while the available information about others is too limited to give a convincing result. Therefore, our target option product will only be 50ETF options from May 2015 to October 2019. We analyse data in monthly frequency since daily data would contain too much mis-pricing information and heavily affect the residual errors of my model. Our dataset includes strike prices, expiration dates, close trading prices and call-put types of all option contracts at the end of each month. We will use them to calculate options implied volatilities as the dependent variable of regression model and details will be described in Section 3.

### **3.3 Stock-relevant Data**

As mentioned before, we would like to exploit sentiment factors to explain variations of option implied volatilities. There are hundreds of sentiment factors argued from various papers and we choose six most representative factors. Neal and Wheatley [7] find discount close-end fund as a crucial factor, so we will test its implication on Chinese markets as well. Share turnover has long been considered as an important technical indicator to capture investor sentiment. Investors normally pay more attention on stock markets and trade more often when they are bullish than bearish. Comparing with secondary markets, investors make decisions more based on their sentiments in primary markets since pre-listing companies have less information and are more difficult to be reasonably priced than listed companies. Baker and Wurgler [8] also argue the number of IPOs and average first day return of IPO companies are higher in a bullish market. The number of new investors would increase sharply as well when the market performs well in the past. Finally, Lemmon and Portniaguina (2006) [10] empirically estimate the strong correlation between consumer confidence and the cross-section of asset prices. Obviously, investor sentiments and expectations based on historical returns are covered in the consumer confidence index. Hence, this level would be our last selected variable.

Overall, we denote DCEF as discount-end fund, TURN as share turnover, IPON as the number of IPOs, IPOR as average first day return of IPO companies, NIA as the new investor amount, and CCI as consumer confidence index in the rest of this paper. One issue we have to encounter is the correlation between different explanatory variables. For example, IPON and IPOR are often positively correlated. New investors participating in markets are associated with higher share turnovers, since naïve investors make investments depended on short-term speculations and would buy and sell stock much more frequently. Therefore, it is necessary to remove the influence of correlations before running regression analysis. Furthermore, we choose consumer price index, industry added value, macroeconomic prosperity index and producer price index as four control variables in robustness tests.

### **3.3 Risk-free Rate**

We choose the historical yield of one-year treasury bonds in China as our risk-free rate. Since short-term bond is more liquid and the returns are more likely to be realized, the option implied volatility approximated by our model will be more accurate corresponding to the maturity.

## **4. Methodology**

### **4.1 Regression Model**

To estimate the relationship between implied volatility and sentiment factors, we formulate a multi-factor regression model. The dependent variable is average options implied volatility at month  $t$ , and the independent variables are sentiment factors mentioned in section 3. We formulate lagged explanatory variables with lags of both one period and two periods to take autocorrelation problem into consideration. However, the correlation and autocorrelation between explanatory variables would influence the accuracy and validation of my model, and that is why we apply principal component analysis before running the tests.

## 4.2 Principal Component Analysis

There are totally twelve factors in our model, with potentially high correlations and autocorrelations with each other. Furthermore, high dimensionality of the variables may result in collinearity problem, which is the limits of traditional regression models. Therefore, we apply machine learning methods to figure these problems out. The principal component analysis, simplified as PCA, serves for reducing variable dimensionality and saving the most significant components without correlation between each other. Therefore, we apply this advanced methodology to handle our data before doing any empirical analysis.

We exploit principal component analysis twice in our model. We firstly run PCA on the twelve raw factors and retain the first six components. Our results show the first six components cumulatively have over 88% contribution ratio, implying they can explain 88% variances of the model. We calculate the weighted average of the first six components and obtain our first aggregate sentiment factor (ASF). We then validate the correlation between the first ASF and twelve raw sentiment factors.

Our logic is for each pair of sentiment factors, we retain the one with higher correlation with the first ASF and leave out the other with lower correlation with the first ASF. After this comparison, we remain six factors in our model. We then use these six raw sentiment factors to run PCA again and retain the first four components. We implement this strategy for the sake of better removing the influence of correlation and autocorrelation.

We find that the first four components can explain over 87% variances of the model. We also report the correlations between these components in Table 1. The results are appealing since they are all nearly uncorrelated with each other, which demonstrates the effectiveness of our methodology. We then calculate the weighted average of the first four components and get our final aggregate sentiment factor, and we now solve the problems of both model collinearity and variable correlation and autocorrelation.

Table 1 Correlations after the second PCA

|   | 1      | 2      | 3      | 4     |
|---|--------|--------|--------|-------|
| 1 | 1.000  | -0.002 | -0.005 | 0.008 |
| 2 | -0.002 | 1.000  | -0.001 | 0.001 |
| 3 | -0.005 | -0.001 | 1.000  | 0.003 |
| 4 | 0.008  | 0.001  | 0.003  | 1.000 |

## 4.3 Option Implied Volatility

We approximate options implied volatilities applying black-scholes formula [3]. At each period, we calculate a different implied volatility for each option contract, and we take the average as our dependent variable. So far, we have accomplished all procedures of data processing and we could thus estimate the relationships.

## 5. Empirical Results

### 5.1 Option Implied Volatility and Aggregate Sentiment Factor

Before demonstrating empirical analysis, we firstly plot the graphs of option implied volatility and final aggregate sentiment factor, with results shown in Figure 1. The blue line indicates aggregate sentiment factors after the second PCA with y-axis at LHS, and the red line indicates option average implied volatilities with y-axis at RHS. It can be clearly observed that these two variables are highly correlated with each other. Thus, we are confident about our empirical results. After running the regression, we get a 0.106 fitted slope with t-statistic of 5.375. The R-squared of

this univariate regression model is 36.6%. Therefore, the aggregate sentiment factor is positively correlated with and have statistically significant predictability of option implied volatilities.

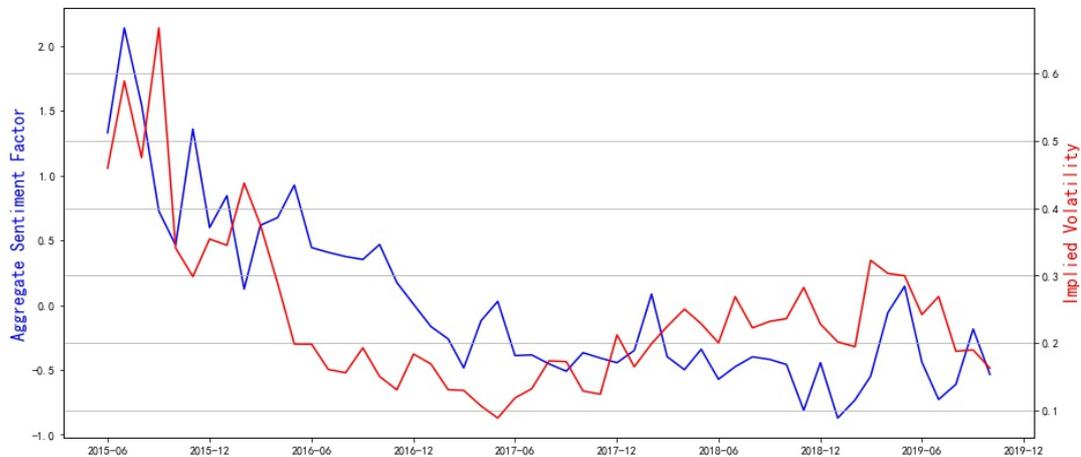


Figure 1 Portfolio average return differences based on maximum daily return over past month.

### 5.2 Option Implied Volatility and Principal Component

In this subsection we run a multi-variable regression analysis and test the implication of each of the four components received from the second PCA on implied volatility. The results are illustrated in Table 2. The coefficients of the first two principal components are positively correlated with dependent variable and are statistically significant. The model R-squared increases to nearly 60%.

Table 2 Statistical Results of PCA

|               | coefficient | t value |
|---------------|-------------|---------|
| 1st component | 0.034       | 4.843   |
| 2nd component | 0.062       | 6.605   |
| 3rd component | -0.011      | -1.082  |
| 4th component | 0.014       | 1.016   |
| R squared     | 59.6%       |         |

### 5.3 Robustness Tests

The results in subsection A and B support the effectiveness of our methodology. However, we would also like to test whether aggregate sentiment factor is still effective after controlling several fundamental macroeconomic variables. Therefore, we add consumer price index, industry added value, macroeconomic prosperity index and producer price index into our model as control variables and run another multi-variable regression as robustness test. The results in Table 3 demonstrate the aggregate sentiment factor is still statistically significant at 95% confidence level, which proves its robustness.

Table 3 Statistical Results of ASF and Control Variables

|           | coefficient | t value |
|-----------|-------------|---------|
| CPI       | 0.008       | 0.678   |
| IVA       | -0.003      | -0.265  |
| MCI       | -0.089      | -3.432  |
| PPI       | 0.038       | 1.251   |
| ASF       | 0.071       | 2.756   |
| R squared | 57.80%      |         |

## 6. Implications

Our works have two main contributions for both academics and industry. Firstly, our model successfully captures the investor sentiment information by linking a bridge between stocks market and options markets. Hence, derivative traders and fund managers could use our aggregate sentiment factors or other similar methods to predict option implied volatility and build investment strategies based on the predictability. We can use implied volatility to estimate stock market sentiment and subsequent performance as well. Secondly, we use more advanced machine learning methodologies to solve problems which are difficult to be encountered by traditional financial models. Therefore, our results also verify the effectiveness and efficiency of machine learning methods in financial analysis and provide researchers with an interesting and meaningful research direction to explore.

## 7. Conclusions

Inspired by ample research about the influence of investor sentiment on asset returns, we focus on testing the correlations and predictability power of sentiment factors obtained from stock markets on option implied volatility. Different from previous works, we use novel methodologies to fix the problems of collinearity and correlations when fitting our results. The principal component analysis is extremely appropriate for dimensionality reduction and the aggregate sentiment factor gained by PCA is validated to strongly and statistically significantly correlated with option average implied volatility. Our results are robust when adding macroeconomic factors as control variables. With this understanding of close relation between stock markets and options, we could analyse investor sentiment and make investment decisions more thoroughly and efficiently. Furthermore, the application of machine learning on finance has been proved to be extraordinary and is a potentially valuable topic for further research.

## References

- [1] Fama E. F. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 1970, 25(2): 383-417.
- [2] Fama E. F., & French, K. R. The cross-section of expected stock returns. *the Journal of Finance*, 1992, 47(2): 427-465.
- [3] Black F., Scholes M. The pricing of options and corporate liabilities. *Journal of political economy*, 1973, 81(3): 637-654.
- [4] Barberis N., Thaler R. A survey of behavioral finance. *Handbook of the Economics of Finance*, 2003, 1: 1053-1128.
- [5] De Bondt W. F., Thaler R. Does the stock market overreact?. *The Journal of finance*, 1985, 40(3): 793-805.
- [6] Barberis N., Shleifer A., Vishny R. A model of investor sentiment. *Journal of financial economics*, 1998, 49(3): 307-343.
- [7] Neal R., Wheatley S. M. Do measures of investor sentiment predict returns?. *Journal of Financial and Quantitative Analysis*, 1998: 523-547.
- [8] Baker M., Wurgler J. Investor sentiment and the cross-section of stock returns. *The journal of Finance*, 2006, 61(4): 1645-1680.
- [9] Han B. Investor sentiment and option prices. *The Review of Financial Studies*, 2008, 21(1): 387-414.
- [10] Lemmon M., Portniaguina E. Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, 2006, 19(4): 1499-1529.