

Optimization of Measuring Investor Sentiment

¹Chunpeng Yang, ²Wei Yan, ³Hanchao Liang

¹*Financial and Securities Research Center, School of Economics and Commerce,
South China University of Technology,
e-mail: chpyang@scut.edu.cn*

²*School of Economics and Commerce, South China University of Technology,
e-mail: wei_yim@163.com*

³*School of Economics and Commerce, South China University of Technology,
e-mail: lianghanchao@126.com*

Abstract

The procedure of selecting investor sentiment proxies scientifically is subjective, and has not been approved by scholars unanimously. This paper proposes an optimization procedure to measure investor sentiment. It strengthens the objectivity and rationality for constructing a sentiment index. Subsequently, this paper takes Chinese stock market as an example, and the result by using the optimization procedure shows that the sentiment index has nice capability to forecast the end of a bear or a bull market. Therefore, the optimization of measuring investor sentiment is efficient.

Keywords: *Investor Sentiment, Optimization Procedure, Sentiment Proxies*

1. Introduction

Behavioral finance has obtained more and more supports since 1980s due to its strong explanation power to financial anomalies which can't be illuminated by the classical finance, and it has become an efficient supplement for classical financial theories. Until 1990s, behavioral finance prospered and became a research category which had integrated system and various branches. More and more scholars made great contributions to this field. During these years from its appearance to prosperity, the research achievements, especially based on investor bias and based on investor sentiment, have shocked the authority of classical financial theories. In the respect of investor bias, Shefrin, Statman [1,2] proposed two famous models according to behavioral finance theory, namely, behavioral asset pricing model (BAPM) and behavioral portfolio theory (BPT), which laid a solid foundation for the research on behavioral finance theory. Hence, many scholars proposed some capital asset pricing models based on investor bias, such as BSV model (Barberis, Shleifer, Vishny [3]), DHS model (Daniel, Hirshleifer, Subrahmanyam [4]), HS model (Hong, Stein [5]) and BHS model (Barberis, Huang, Santos [6]). However, these models based on investor bias have two common shortcomings. First, all of them lacked generality because they aimed to explain a special anomaly of short-term momentum and long-term reversal based on different biases. Any research based on a special bias was unilateral since the great amount of investor biases when they make decision. Therefore, it needs to modify the separate research approach. All investor biases influence their sentiment ultimately. Investor sentiment, as an important branch of behavioral finance, has obtained a great plentiful of achievements. Maybe adopting the sentiment as a method for finance investigation is a nice way. Second, the information of investor bias is difficult to be identified and can't be measured, consequently can't be empirically testified. But the sentiment is easy to be measured by variant methods including investigation questionnaire or financial proxies, especially after Baker, Wurgler [7,8] constructed a composite sentiment index for the aggregate finance market.

Researches about sentiment developed the studies on some traditional financial problems, such as asset's pricing (Yang, Yan [9]) and portfolio selection (Yang, Xie [10]). Although there has been about 20 years for the study on investor sentiment, the research in this category is still not profound due to the difficult measure of sentiment index. Nowadays, almost all researches are empirical tests for the relation between sentiment and market performance. As said by Baker, Wurgler [8], most of the early researches focused on whether the sentiment affects asset price or returns. After the consensus by scholars, the key problem is how to measure the sentiment index scientifically. Baker, Wurgler [8] also pointed out that there were two approaches for the research on investor sentiment. One approach is

“bottom up” which means the researches mainly aim to interpret the effect mechanism of every psychological bias on the sentiment and consequently to study investor’s abnormal behaviors and market anomalies. The other approach is “top down” which implies to trace the investor sentiment’s influence on the aggregate market or individual stock returns. Furthermore, the basic premise of the latter approach is to measure the sentiment quantificationally. For the historical methods of measure and description to investor sentiment, Brown, Cliff [11] and Yan, Yang, Xie [12] have summarized. They pointed out that the current researches tend to employ multiple proxy variables to describe the high or low aggregate market sentiment, and then quantify the sentiment indirectly and synthetically to study the interaction between the market performance and sentiment.

The most prominent achievement in measuring the sentiment by using multiple proxies is Baker, Wurgler [7,8] who constructed an indirect market sentiment index referred to BW index by scholars later. They selected six variables, such as closed-end fund discount, turnover rate, number of IPO, first-day return of IPO, equity share in new issues, dividends, as the proxies of market aggregate sentiment. Liao, Huang, Wu [13] also chose ten variables as the proxies of investor sentiment which are: individual stock return, individual stock trading volume, S&P500 index return, Russell 2000 index return, S&P500 index trading volume, S&P500 index option put/call ratio variation, first-day return of IPO, number of IPO, NYSE share turnover, and net purchases of mutual funds. Finter, Niessen-Ruenzi, Ruenzi [14] employed Germany’s GfK consumer confidence, trading volume, net fund flows, IPO returns, number of IPOs, equity/debt ratio, put-call ratio as sentiment proxies. Kim, Ha [15] used Korea’s ten variables which are the equity share in new issues, relative strength index, money flow index, SERI financial strength index, the average first-day returns on IPOs, the consumer’s deposit for stock investment, SERI consumer sentiment index, psychological line, margin borrowing, and KOSPI share turnover.

In a word, the indirect and composite sentiment index has become the most commonly used indexes of sentiment. Its forecast power to market performance and the ability of identification for cross-section effect are approved unanimously by scholars. However, the selection of sentiment proxies in the procedure of measuring an indirect and composite sentiment index is subjective and illogical, including variables of market performance, variables of market structure, variables of special products such as the funds or the initial public offerings, and even the variables of derivatives. The methods used by previous studies about how to select the proxies more scientifically have not been approved by scholars unanimously. There will be some statistical errors if the selection of sentiment proxies is subjective and illogical. Also it is not conducive to the scientific framework for the research on investor sentiment. Hence, it is necessary to discuss how to select the sentiment proxies more objectively, which will be the kernel content studied in this paper.

The purpose of this paper is to propose an optimization procedure for selecting the proxies to construct the investor sentiment index, especially of the aggregate market sentiment index. This paper will improve the early methods of selecting sentiment proxies, and it will weaken their subjectivity and arbitrariness for measuring the sentiment while strengthen the objectivity and rationality, which is the most significant contribution in this paper.

The rest of this paper is arranged as follows: section 2 proposes the optimization procedure for measuring the sentiment; section 3 presents an example by using the method in section 2 and the data in Chinese stock market; section 4 gives the total conclusions of this paper.

2. Optimization procedure for measuring investor sentiment

Proxies selecting is miscellaneous and subjective, besides, there is no reasonably verifiable introduction and no rejection to unreasonable proxies, so there is no scientific procedure and basis for the constructing of sentiment index. To solve this problem, it’s necessary to choose the right proxies with some standard procedure and appraisable criterion. In the following part, we propose an optimization procedure including the standard procedure and appraisable criterion for choosing sentiment proxies, which may be helpful for future research on the constructing of investor sentiment index. According to this rule and procedure, we will construct a more reasonable sentiment index consequently.

There are four steps to measure the investor sentiment:

Step 1: select the original proxies and gather their data

The sentiment can be divided into institutional investor sentiment, individual investor sentiment, as well as the aggregate market sentiment, according to the sentiment carrier. Different investor sentiment needs different proxies. So, we should firstly ascertain the scope of proxies from the influence mechanism, that is, make sure that which variables affect this type of investor sentiment. There are three problems about ascertaining the scope of sentiment proxies in current researches, which need to be modified:

First, the mixed use of level variables and rate variables. Usually, there is a range for rate variable, from 0 to 1 for instance, which is suitable to describe the fluctuation of asset price, meanwhile there is no limit to the level variable which is referred to describe the trend of asset price. For example, at the end of bull market, the composite index often climbs to a new high value and the aggregate first day returns of IPO get new high in the situation of larger issues of IPO, but the first day return rate of IPO is not higher than before sometimes. By all means, the sentiment index constructed by mixed use of level variables and rate variables will be counteracted, and its values will deviate from some real values, and perhaps can't reflect neither the fluctuation of asset price nor the trend of it. So, all of the sentiment proxies should be level if the research purpose is to study the trend of asset price, and be corresponding rate variables in order to investigate the fluctuation or return rate of asset.

Second, the neglect for an important type of variable. Many early researches results imply that asset price or return rate one period ago has significant effects on the investor sentiment. But few of them neither used these conclusions, nor chose the asset price or return rate one period ago as sentiment proxy. Therefore, the future research should select one of them as a proxy.

Third, the neglect for self-influence of variable. The current sentiment is not only affected by current values of proxies, but also by the sentiment itself one period ago. So, it is necessary to study the possibility of the proxies' values one period ago be as sentiment proxies, which has not been seen in the previous researches. Now, take every proxy's values one period ago as an added sentiment proxy. Then, the number of proxies becomes two times than that before. All the proxies at this time are treated as the original proxies for sentiment.

After completing the former procedures, the data of all original proxies should be gathered in order to analyze quantitatively. The following steps are given to solve the problems whether original proxies can describe investor sentiment, which of them are better, and how to eliminate the unreasonable proxies.

Step 2: Correlation analysis between every original proxy and asset performance

Now the question is that which original proxies are more appropriate. In order to solve this problem, the correlation between asset price or return rate and every original proxy should be analyzed. Some most appropriate proxies will be reserved after the elimination of unreasonable original proxies through the correlation analysis introduced as below.

It is apparent that the stronger of the correlation between asset performance and an original proxy, the better for this proxy to describe the investor sentiment. It will reduce the quality of sentiment index if we select the proxies which have weak correlations with asset performance into the sentiment proxies, although these proxies may have some relationship with investor sentiment. In the correlation analysis process, the analysis should be done between asset price and every original proxy which is level, such as the stock price and trading volume, while be done between asset return rate and every original proxy which is rate variable, such as stock return rate and turnover. The coefficients are obtained after the correlation analysis. Suppose the correlation coefficients between every original proxy and asset performance are ρ_i , and the mean of all coefficients' absolute values is μ , standard deviation is σ . If adopting the boundary value $\mu - \sigma$ or some else, the original proxy satisfied $|\rho_i| < \mu - \sigma$ should be eliminated. Take the residual proxies as the selected proxies.

Step 3: Correlation analysis among all of the selected proxies

If the correlations between some sentiment proxies are too strong, the effect from a certain direction will be emphasized, and the sentiment values will deviate from some real values. Hence, the correlations between the selected proxies each other should be analyzed in order to eliminate some variables of low quality. Suppose the coefficients are $\rho_{ij}, (i \neq j)$, and the mean of all coefficients' absolute values is e , standard deviation is ε . If adopting the boundary value $e + \varepsilon$ or some else, one of the two selected proxies whose coefficients' absolute value is bigger than the boundary value $\rho_{ij} > e + \varepsilon$ should be taken into account to eliminate. The rule is to eliminate one selected proxy

(i or j) whose coefficient's absolute value with asset performance is smaller, which has been done in step 2. Now, take the residual proxies as the final proxies to construct the investor sentiment index.

Step 4: The standardization of final proxies and construction of investor sentiment index

The final proxies should be standardized in order to make them equivalent and comparable. Then the investor sentiment index will be obtained by a certain mathematical approach. Generally, the standardization of the proxies often changes them to be 0-mean and 1-variance, and the usual approach is principal component analysis.

3. An example for measuring the aggregate market sentiment

For the capital asset pricing problem and the portfolio selection problem based on investor sentiment, the most important task is to construct an individual stock's sentiment index, and then investigate the influences on asset's price or investment weight. However, there are a lot of plentiful research achievements on aggregate market, institutional investor and individual investor, but few of them are about the study on sentiment index for a special individual stock. There are mainly two methods for individual stock so far: (1) Kaniel, Saar, Titman [16] proposed a sentiment index based on the trading volume of individual account; (2) Frazzini, Lamont [17] proposed another one based on the mutual fund flows of an individual stock. Nevertheless, both of these methods adopted only one proxy variable to be the sentiment index, which can't describe the sentiment characteristics for an individual stock entirely. Therefore, we should consider adopting several proxies to describe the sentiment for an individual stock systematically. Considering that most of the current studies on sentiment constructed the aggregate market sentiment, this paper will still construct the aggregate market sentiment index.

As mentioned in section 1, proxies selecting for the aggregate market sentiment is miscellaneous and subjective. Compared with the previous studies, this paper will choose some more scientific and appropriate proxies, by using the optimization procedures including the standard procedure and appraisable criterion proposed in section 2. At the end, we will get an aggregate market sentiment index.

This paper will take the Chinese stock market, especially of Shanghai composite index, as example.

Step 1: select the original proxies and gather their data

Considered the previous literatures, comprehensively with the data characteristics of Chinese stock market, this paper intends to select the sentiment proxies such as the current new-open stock account of Shanghai market NS_t , current new-open fund account of Shanghai market NF_t , current trading volume of Shanghai composite index TV_t , current Shanghai fund index PF_t , Shanghai composite index lagged one period PS_{t-1} . In addition, we'll consider that whether the lagged one period value of all proxies may be the sentiment proxies, eventually, this paper selects additional variables NS_{t-1} , NF_{t-1} , TV_{t-1} , PF_{t-1} , PS_{t-2} , altogether 10 proxies, as original proxies. The proxies are all weekly data from Feb. 6, 2004 to Apr. 30, 2008.

The current literature, just like Baker, Wurgler [7,8], and so on, adopted some variables as closed-end fund discount, number of IPO and first-day return of IPO. However, this paper doesn't employ them due to the characteristics of Chinese stock market. The number of closed-end fund too small to describe the properties and information about funds, which maybe distort the quality of sentiment. Meanwhile, IPO process has been intermitted for several times, and the data were absent for a long period. Therefore, the variables including closed-end fund discount, number of IPO and first-day return of IPO are not suitable for the description of Chinese aggregate market sentiment.

Step 2: Correlation analysis between every original proxy and asset performance

Now we analyze the correlation between every original proxy and the current value of Shanghai composite index, the coefficients are shown in Table 1. The mean of all coefficients' absolute values in Table 1 is $\mu=0.8689$, standard deviation is $\sigma=0.1384$. If adopting the boundary value $\mu-\sigma=0.7305$, the original proxies NF_t and NF_{t-1} should be eliminated because the absolute value of their coefficients with Shanghai composite index is smaller than the boundary value. It means that, compared with other proxies, current and lagged one period new-open fund account of Shanghai market are not suitable to be the market sentimental proxies due to the small correlations with Shanghai

composite index. So, treat the other eight proxies as selected proxies.

Table 1. Coefficients between every original proxy and the asset performance

	NS_t	NF_t^*	TV_t	PF_t^{**}	PS_{t-1}	NS_{t-1}^{**}	NF_{t-1}^*	TV_{t-1}	PF_{t-1}^{**}	PS_{t-2}^{**}
PS_t	0.8231	0.6198	0.8640	0.9734	0.9963	0.8190	0.6330	0.8694	0.9683	0.9919

Remark: Superscript * means the proxies eliminated in step 2, and ** means the proxies eliminated in step 3

Step 3: Correlation analysis among all of the selected proxies

Now we analyze the correlation among the selected proxies, the coefficients are shown in Table 2. The mean of all coefficients' absolute values in Table 2 is $\mu' = 0.8674$, standard deviation is $\sigma' = 0.0852$. If adopting the boundary value $\mu + \sigma = 0.9527$, one of the two selected proxies whose coefficients' absolute value is bigger than the boundary value should be eliminated. The rule is to eliminate the proxy whose coefficients' absolute value with Shanghai composite index is smaller, which has been done in step 2. According to this, PF_t , NS_{t-1} , PF_{t-1} , PS_{t-2} should be eliminated, and the final proxies are NS_t , TV_t , PS_{t-1} , TV_{t-1} . It means that most information of the eliminated proxies is included in the final proxies, and it doesn't need to measure again. Otherwise, there will be a deviation for the sentiment index from some real values.

Table 2. Coefficients among the selected proxies

	NS_t	TV_t	PF_t	PS_{t-1}	NS_{t-1}	TV_{t-1}	PF_{t-1}
TV_t	0.9124						
PF_t	0.7359	0.8240					
PS_{t-1}	0.8095	0.8445	0.9745*				
NS_{t-1}	0.9705*	0.8939	0.7457	0.8209			
TV_{t-1}	0.9098	0.9365	0.8259	0.8650	0.9203		
PF_{t-1}	0.7281	0.8071	0.9969*	0.9751*	0.7421	0.8211	
PS_{t-2}	0.7939	0.8292	0.9762*	0.9963*	0.8110	0.8457	0.9767*

Remark: Coefficient corresponding to two proxies with superscript * satisfies the eliminated rule

Step 4: The standardization of final proxies and construction of investor sentiment index

Standardize the final proxies NS_t , TV_t , PS_{t-1} and TV_{t-1} , the method is that every proxy minus its mean value then divide its standard deviation. By employing the method of principal component analysis, we take the first principal component as sentiment index $Sent$. Therefore, the relation between sentiment index and every final proxy is:

$$Sent = 0.5049 \cdot NS_t + 0.4840 \cdot TV_{t-1} + 0.4993 \cdot PS_{t-1} + 0.5114 \cdot TV_t$$

Thus, the eigenvalues of final proxies' covariance matrix are 3.6420, 0.1956, 0.1056 and 0.0568, respectively. It implies that the sentiment index $Sent$ includes 91.05% information of these four final proxies.

The Shanghai composite index and the aggregate market sentiment index are shown in Figure 1.

The solid line in Figure 1 is the aggregate market sentiment index whose values are scaled by the right y-axis, and the dotted line is the actual weekly closing price of the Shanghai composite index whose values are scaled by the left y-axis. The maximum of sentiment is 5.6811, and the minimum is -1.5226.

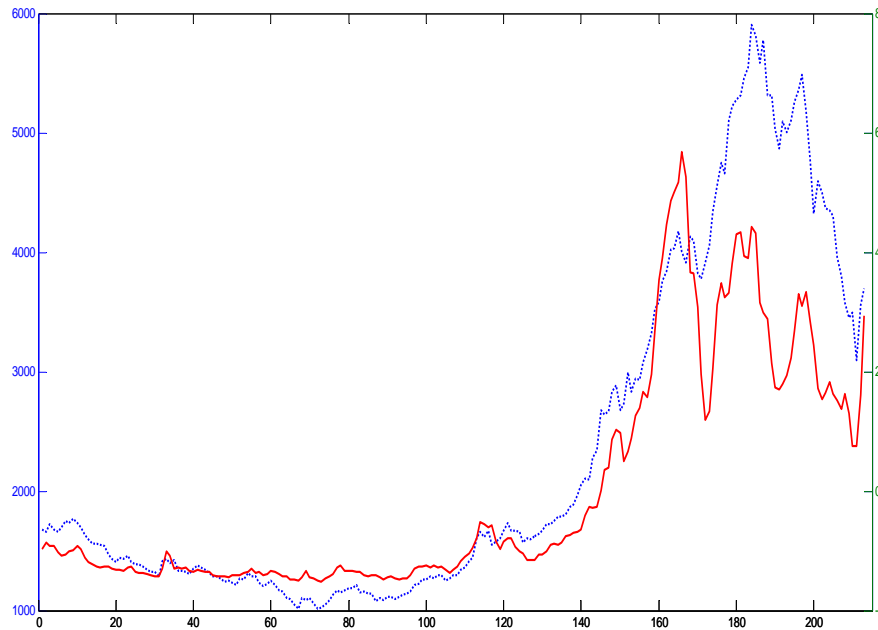


Figure 1. Shanghai composite index and the aggregate market sentiment index

According to the history data, it can be seen that there are one bull market and one bear market. At the end of the bear market, the Shanghai composite index dropped to a new low value from the 60th week to 70th week, that is, Apr. 22, 2005 to Jul. 8, 2005. But the sentiment index didn't get a new low in this period. This bottom deviation implied that the new low price of Shanghai composite index was not agreed by all of the investors in the aggregate market. At the end of the bull market, the Shanghai composite index climbed to a new high value from the 168th week to 181st week, that is, Jul. 6, 2007 to Oct. 12, 2007. But the sentiment index didn't get a new high in this period. This top deviation implied that the new high price of Shanghai composite index was not supported by all of the investors in the aggregate market. Maybe the relationship between sentiment index and Shanghai composite index does forecast the end of a bear or a bull market efficiently in some meanings.

4. Conclusion

Although the research on investor sentiment has been nearly 20 years, it still just focuses on the empirical study. The basic premise of current investigations is to measure the sentiment efficiently and quantificationally. Nowadays, the current researches tend to employ multiple proxy variables to describe the aggregate market sentiment, and the indirect and composite sentiment index has become the most commonly used indexes of sentiment. However, the selection of sentiment proxies in the procedure of measuring an indirect and composite sentiment index is subjective and illogical. The methods used by previous studies about how to select the proxies more scientifically have not been approved by scholars unanimously. Hence, it needs an optimization procedure to measure the sentiment, which has been done by this paper. Subsequently, an example of Chinese stock market has shown that the sentiment index has nice capability of forecast towards the end of a bear or a bull market. The optimization of measuring the investor sentiment is efficient.

Future researches, especially of theoretical researches, can be studied on the possibility of sentiment constructed by using the optimization procedure proposed in this paper as a factor of pricing for the asset, based on the three-factors pricing model (Fama, French [18]), the four-factors pricing model (Carhart [19]) and the four-factors pricing model (Kumar, Lee [20]), and so on. Moreover, the relationship between this sentiment index and portfolio problem should be investigated, too.

5. Acknowledgments

The authors gratefully acknowledge the support by “National Natural Science Foundation of China (70871042)”.

6. References

- [1] Shefrin H, Statman M, “Behavioral capital asset pricing theory”, *Journal of Financial and Quantitative Analysis*, vol. 29, no. 3, pp. 323-349, 1994.
- [2] Shefrin H, Statman M, “Behavioral portfolio theory”, *Journal of Financial and Quantitative Analysis*, vol. 35, no. 2, pp. 127-151, 2000.
- [3] Barberis N, Shleifer A, Vishny R, “A model of investor sentiment”, *Journal of Financial Economics*, vol. 49, no. 3, pp. 307-343, 1998.
- [4] Daniel K, Hirshleifer D, Subrahmanyam A, “Investor psychology and security market under and overreactions”, *Journal of Finance*, vol. 53, no. 6, pp. 1839-1885, 1998.
- [5] Hong H, Stein J, “A unified theory of underreaction, momentum trading, and overreaction in asset markets”, *Journal of Finance*, vol. 54, no. 6, pp. 2143-2184, 1999.
- [6] Barberis N, Huang M, Santos T, “Prospect theory and asset prices”, *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 1-53, 2001.
- [7] Baker M, Wurgler J, “Investor sentiment and the cross-section of stock returns”, *Journal of Finance*, vol. 61, no. 4, pp. 1645-1680, 2006.
- [8] Baker M, Wurgler J, “Investor sentiment in the stock market”, *Journal of Economic Perspective*, vol. 21, no. 2, pp. 129-151, 2007.
- [9] Yang C, Yan W, “Does high sentiment cause negative excess return”, *Journal of Digital Content Technology and its Applications*, vol. 5, no. 12, pp. 211- 217, 2011.
- [10] Yang C, Xie J, “Sentiment perceived portfolio optimization”, *Journal of Convergence Information Technology*, vol. 6, no. 12, pp. 203-209, 2011.
- [11] Brown G, Cliff M. “Investor sentiment and asset valuation”, *Journal of Business*, vol. 78, no. 2, pp. 405-440, 2005.
- [12] Yan W, Yang C, Xie J, “A survey for researches on investor sentiment in the financial Market”, *Journal of South China University of Technology (Social Science Edition)*, vol. 13, no. 3, pp. 33-43, 2011. (in Chinese)
- [13] Liao T, Huang C, Wu C, “Do fund managers herd to counter investor sentiment?”, *Journal of Business Research*, vol. 64, no. 2, pp. 207-212, 2011.
- [14] Finter P, Niessen-Ruenzi A, Ruenzi S, “The impact of investor sentiment on the German stock market”, SSRN Working Paper, 2010.
- [15] Kim T, Ha N, “Investor sentiment and market anomalies”, SSRN Working Paper, 2010.
- [16] Kaniel R, Saar G, Titman S, “Individual investor sentiment and stock returns”, New York University, Working Paper, 2004.
- [17] Frazzini A, Lamont O, “Dumb money: mutual fund flows and the cross-section of stock returns”, *Journal of Financial Economics*, vol. 88, no. 2, pp. 299-322, 2008.
- [18] Fama E, French K, “Common risk factors in returns on stocks and bonds”, *Journal of Financial Economics*, vol. 33, no. 1, pp. 3-56, 1993.
- [19] Carhart M, “On persistence in mutual fund performance”, *Journal of Finance*, vol. 52, no. 1, pp. 57-82, 1997.
- [20] Kumar A, Lee C, “Retail investor sentiment and return comovements”, *Journal of Finance*, vol. 61, no. 5, pp. 2451-2486, 2006.